THE ANNALS
of
MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XI, No. 4 — December, 1940

THE ANNALS OF MATHEMATICAL STATISTICS

EDITED BY

S. S. WILKS, *Editor*

A. T. CRAIG J. NEYMAN

WITH THE COÖPERATION OF

H. C. CARVER R. A. FISHER R. DE MISES
H. CRAMÉR T. C. FRY E. S. PEARSON
W. E. DEMING H. HOTELLING H. L. RIETZ
G. DARMOIS W. A. SHEWHART

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, P. R. Rider, Washington University, St. Louis, Mo.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $4.00 per year. Single copies $1.25. Back numbers are available at the following rates:

Vols. I-IV $5.00 each. Single numbers $1.50.
Vols. V to date $4.00 each. Single numbers $1.25.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879

he
re,
n-
It.
ti-
St.

cs
ts
py
er
ot-
on
to
ies

ut
st.

25.

THE DISTRIBUTION THEORY OF RUNS

By A. M. Mood

1. Introduction. In studying a particular sample, the order in which the elements of the sample were drawn is frequently available to the statistician. This important information is usually entirely neglected by him. Such disregard must be attributed, to a considerable extent, to the unsatisfactory state of mathematical devices for using the knowledge in question. One reasonable mathematical method for handling this information, the one to be used in this paper, is to make use of the distribution of runs. A run is defined as a succession of similar events preceeded and succeeded by different events; the number of elements in a run will be referred to as its length.

The distribution theory of runs has had a stormy career. The theory seems to have been started toward the end of the nineteenth century rather than in the days of Laplace when there was so much interest in games of chance. In 1897 Karl Pearson [1], in a discussion of data taken from the roulette tables at Monte Carlo, wrote "... the theory of runs is a very simple one." In this book he developed no theory but it is evident from his computations that he regarded the distribution of runs as a special case of the multinomial distribution. The multinomial method, besides evading the issue somewhat and raising questions of random sampling, also gives incorrect results when one is interested in runs of more than one kind of element. In 1899 Karl Marbe [2] derived an expression for the mean of the number of iterations of a given length from a binomial population. This result was incorrect because he neglected dependence between overlapping iterations. An iteration is defined as a sequence of similar events; a run of length t is counted as $t - s + 1$ iterations of length s for $s \leq t$. Marbe has assembled a great mass of data with the object of proving the popular hypothesis that a "head" becomes highly probable after a long succession of "tails" has appeared. Ordinary significance tests applied to his data do not support this contention, but Marbe continues to advocate it [3] and [5]. Of course, he has been severely criticised by many mathematical statisticians.

In 1904 Grünbaum [6] derived the mean of the number of runs of given length from a binomial population by the multinomial method. The first correct formulae were derived in 1906 by Bruns [7] who found the mean and variance of the number of iterations of given length in samples from a binomial population. In a book published in 1917 von Bortkiewicz correctly derived for the first time the mean and variance of runs from a binomial population using a method similar to that of Bruns. This book [8] contains a great many formulae for means and variances of runs and iterations under various special circumstances; a large portion of it is devoted to an exhaustive criticism of Marbe's work. In 1921 von

Mises [9] showed that the number of long runs of given length was approximately distributed according to the Poisson law for large samples.

It was not until 1925 (so far as the author has been able to ascertain) that an actual distribution function appeared when Ising [10] gave the number of ways of obtaining a given total number of runs (without regard to length) from arrangements of two kinds of elements. Stevens [12] in 1939 published the same distribution and described a χ^2 criterion for significance. Wald and Wolfowitz [13] in 1940 published the same distribution and showed that it was asymptotically normal. These papers are all concerned with random arrangements of a fixed number of elements of each of two kinds; the last mentioned paper describes a very interesting application of the distribution to the problem of testing the hypothesis that two samples have come from the same continuous distribution. Wishart and Hirshfeld [11] in 1936 derived the distribution of the total number of runs (again without regard to length) in samples from a binomial population and showed it was asymptotically normal.

In this paper we shall derive distributions of runs of given length both from random arrangements of fixed numbers of elements of two or more kinds, and from binomial and multinomial populations. Also we shall give the limiting form of these distributions as the sample size increases. These limiting distributions are all normal. The distribution problem is, of course, a combinatorial one, and the whole development depends on some identities in combinatory analysis,—some new and some well known to students of partition theory.

The paper will be divided into two parts. The first will deal with distributions obtained from random arrangements of a fixed number of each kind of element. The second will deal with distributions of elements from a binomial or multinomial population.

Part I

2. Distribution of runs of two kinds of elements. Consider random arrangements of n elements of two kinds, for example n_1 a's and n_2 b's with $n_1 + n_2 = n$. Let r_{1i} denote the number of runs of a's of length i, and let r_{2i} denote the number of runs of b's of length i. For example the arrangement

$$a\ b\ b\ a\ b\ a\ a\ a\ b\ b\ a\ a\ a$$

will be characterized by the numbers $r_{11} = 2$, $r_{13} = 2$, $r_{21} = 1$, $r_{22} = 2$, and all other $r_{ij} = 0$. Also we let $r_1 = \sum_i r_{1i}$ and $r_2 = \sum_i r_{2i}$ denote the total number of runs of a's and b's respectively. Throughout the paper a binomial coefficient will be denoted by

$$\binom{m}{k} = \frac{m!}{k!(m-k)!} \tag{2.1}$$

and this is defined to be zero when $m < k$. A multinomial coefficient will often be denoted by

$$\left[\begin{matrix} m \\ m_i \end{matrix}\right] = \frac{m!}{m_1!m_2!\cdots m_s!} \tag{2.2}$$

$$\Sigma m_i = m, \qquad m_i \geq 0 \tag{2.3}$$

and when such a coefficient is to be summed over the indices m_i the two conditions (2.3) are always understood and will not be repeated; other conditions on the indices will be placed below the summation sign.

Given a set of numbers r_{ij} ($i = 1, 2; j = 1, 2, \cdots, n_i$) such that $\sum_j jr_{ij} = n_i$, there are $\left[\begin{matrix} r_1 \\ r_{1j} \end{matrix}\right]$ and $\left[\begin{matrix} r_2 \\ r_{2j} \end{matrix}\right]$ different arrangements of the runs of a's and b's respectively. Hence the total number of ways of obtaining the set r_{ij} is

$$N(r_{ij}) = \left[\begin{matrix} r_1 \\ r_{1j} \end{matrix}\right]\left[\begin{matrix} r_2 \\ r_{2j} \end{matrix}\right] F(r_1, r_2) \tag{2.4}$$

where $F(r_1, r_2)$ is the number of ways of arranging r_1 objects of one kind and r_2 objects of another so that no two adjacent objects are of the same kind. Thus

$$\begin{aligned} F(r_1, r_2) &= 0 \quad \text{if} \quad |r_1 - r_2| > 1, \\ &= 1 \quad \text{if} \quad |r_1 - r_2| = 1, \\ &= 2 \quad \text{if} \quad r_1 = r_2 \quad . \end{aligned} \tag{2.5}$$

Since there are $\binom{n}{n_1}$ possible arrangements of the a's and b's, we have at once the distribution of the r_{ij}

$$P(r_{ij}) = \frac{\left[\begin{matrix} r_1 \\ r_{1j} \end{matrix}\right]\left[\begin{matrix} r_2 \\ r_{2j} \end{matrix}\right] F(r_1, r_2)}{\binom{n}{n_1}}. \tag{2.6}$$

Certain marginal distributions will also be of interest. To obtain, for example, the distribution of the r_{1j}, it is first necessary to sum $\left[\begin{matrix} r_2 \\ r_{2j} \end{matrix}\right]$ over all partitions of n_2. This is easily accomplished by finding the coefficient of x^{n_2} in

$$(x + x^2 + x^3 + \cdots)^{r_2} = x^{r_2}(1 + x + x^2 + \cdots)^{r_2} = \frac{x^{r_2}}{(1-x)^{r_2}}$$
$$= x^{r_2} \sum_{t=0}^{\infty} \binom{r_2 - 1 + t}{r_2 - 1} x^t.$$

The term corresponding to $t = n_2 - r_2$ gives the desired result:

$$\sum_{\Sigma jr_{2j} = n_2} \left[\begin{matrix} r_2 \\ r_{2j} \end{matrix}\right] = \binom{n_2 - 1}{r_2 - 1}. \tag{2.7}$$

We have then

$$(2.8)\qquad P(r_{1j}, r_2) = \frac{\begin{bmatrix} r_1 \\ r_{1j} \end{bmatrix}\binom{n_2 - 1}{r_2 - 1} F(r_1, r_2)}{\binom{n}{n_1}}$$

and summing this over r_2, a slight simplification gives

$$(2.9)\qquad P(r_{1j}) = \frac{\begin{bmatrix} r_1 \\ r_{1j} \end{bmatrix}\binom{n_2 + 1}{r_1}}{\binom{n}{n_1}}.$$

The distribution (2.6) summed over r_{1j} and r_{2j} gives by means of (2.7)

$$(2.10)\qquad P(r_1, r_2) = \frac{\binom{n_1 - 1}{r_1 - 1}\binom{n_2 - 1}{r_2 - 1} F(r_1, r_2)}{\binom{n}{n_1}}$$

which is essentially the distribution derived by Wald and Wolfowitz [13], and summing this over r_2 we get the distribution discussed by Stevens [12]

$$(2.11)\qquad P(r_1) = \frac{\binom{n_1 - 1}{r_1 - 1}\binom{n_2 + 1}{r_1}}{\binom{n}{n_1}}.$$

Another marginal distribution which will be useful is obtained by summing (2.9) over r_{1i} for $i \geq k$. If we let

$$s_{1j} = r_{1j}, \qquad j < k,$$

$$s_{1k} = \sum_{k}^{n_1} r_{1j}, \qquad A = \sum_{1}^{k-1} j r_{1j},$$

we must then sum the multinomial coefficient

$$\frac{s_{1k}!}{r_{1k}! \cdots r_{1n_1}!}$$

over all partitions of $n_1 - A$ such that every part is greater than $k - 1$. This is given by the coefficient of $x^{n_1 - A}$ in

$$(x^k + x^{k+1} + \cdots)^{s_{1k}} = x^{ks_{1k}} \sum_{t=0}^{\infty} \binom{s_{1k} - 1 + t}{s_{1k} - 1} x^t$$

thus we have

$$(2.12)\qquad \sum_{(k)} \frac{s_{1k}!}{r_{1k}! \cdots r_{1n_1}} = \binom{n_1 - A - (k-1)s_{1k} - 1}{s_{1k} - 1}$$

where $\sum_{(k)}$ denotes summation over all positive integers $r_{1k}, r_{1\,k+1}, \cdots, r_{1n_1}$ such that $\sum_{k}^{n_1} jr_{1j} = n_1 - A$. This identity with (2.9) gives

$$(2.13)\qquad P(s_{1i}) = \frac{\begin{bmatrix} s_1 \\ s_{1i} \end{bmatrix} \binom{n_2 + 1}{s_1} \binom{n_1 - A - (k-1)s_{1k} - 1}{s_{1k} - 1}}{\binom{n}{n_1}}, \quad i = 1, 2, \cdots, k.$$

Another useful distribution analogous to (2.13) is derived by considering runs of both kinds of elements. If we define s_{2j} $(j = 1, 2, \cdots, h)$ and B in terms of r_{2j} just as s_{1i} and A were defined above, it follows at once from (2.6) and (2.12) that

$$(2.14)\qquad P(s_{1i}, s_{2j}) =$$

$$\frac{\begin{bmatrix} s_1 \\ s_{1i} \end{bmatrix}\begin{bmatrix} s_2 \\ s_{2j} \end{bmatrix} \binom{n_1 - A - (k-1)s_{1k} - 1}{s_{1k} - 1} \cdot \binom{n_2 - B - (h-1)s_{2h} - 1}{s_{2h} - 1} F(s_1, s_2)}{\binom{n}{n_1}},$$

$$i = 1, 2, \cdots, k; j = 1, 2, \cdots, h.$$

These last two distributions should be the most useful for applications. The long runs have been added together to form the new variables s_{1k} and s_{2h} thus decreasing materially the number of variables as compared with (2.6) and (2.9) while at the same time little information is lost. One is free to choose k and h so that the number of variables is appropriate for the data at hand. Moreover, it is shown in Section 5 that these variables are asymptotically normally distributed so that one may apply a simple χ^2 test of significance for "randomness of elements with respect to order" when dealing with large samples. We shall then be able to test whether a sample has been "randomly" drawn in a certain sense.

3. Moments for runs of two kinds of elements. Instead of dealing with the ordinary moments we shall obtain formulae for the factorial moments because the expressions are much more compact. As is customary, a factorial will be denoted by

$$(3.1)\qquad x^{(a)} = x(x-1)(x-2)\cdots(x-a+1),$$

and $x^{(0)}$ is defined to be 1. Of course the ordinary moments are determined by the factorial moments by means of relations of the type

$$x^a = \sum_{i=0}^{a} C_i^a x^{(i)}.$$

A recent discussion of the coefficients C_i^a has been given by Joseph [14]. The mathematical expectation of a function $f(r)$ will be denoted by

$$E(f(r)) = \sum_r f(r)P(r). \tag{3.2}$$

Of course E is a linear operator. We shall require the following identity

$$\sum_{(1)} \prod_i r_{1i}^{(a_i)} \begin{bmatrix} r_1 \\ r_{1i} \end{bmatrix} = r_1^{(\Sigma a_i)} \binom{n_1 - \sum i a_i - 1}{r_1 - \sum a_i - 1} \tag{3.3}$$

where $\Sigma_{(1)}$ denotes summation over all positive integers $r_{11}, r_{12}, \cdots, r_{1n_1}$ such that $\sum_1^{n_1} ir_{1i} = n_1$. (3.3) may be verified by differentiating

$$\varphi(t_i) = (t_1 x + t_2 x^2 + \cdots)^{r_1}$$

a_i times with respect to t_i $(i = 1, 2, \cdots, n_1)$, then finding the coefficient of x^{n_1} after putting $t_i = 1$. The identity (3.3) enables us to find the factorial moments of the variables in the distribution (2.9) for we have

$$\begin{aligned} E\left(\prod_i r_{1i}^{(a_i)}\right) &= \sum_{r_{1i}} \prod r_{1i}^{(a_i)} \begin{bmatrix} r_1 \\ r_{1i} \end{bmatrix} \binom{n_2+1}{r_1} \bigg/ \binom{n}{n_1} \\ &= \sum r_1^{(\Sigma a_i)} \binom{n_1 - \sum i a_i - 1}{r_1 - \sum a_i - 1} \binom{n_2+1}{r_1} \bigg/ \binom{n}{n_1} \\ &= \sum (n_2+1)^{(\Sigma a_i)} \binom{n_1 - \sum i a_i - 1}{r_1 - \sum a_i - 1} \cdot \binom{n_2 - \sum a_i + 1}{r_1 - \sum a_i} \bigg/ \binom{n}{n_1} \qquad (3.4) \\ &= (n_2+1)^{(\Sigma a_i)} \binom{n - \sum (i+1)a_i}{n_1 - \sum i a_i} \bigg/ \binom{n}{n_1}. \end{aligned}$$

The sum on r_1 involved in the last step is given by the identity

$$\sum_{i=0}^{B} \binom{A}{C+i} \binom{B}{i} = \binom{A+B}{C+B} \tag{3.5}$$

which is readily obtained by equating coefficients of x^C in

$$(1+x)^A \left(1 + \frac{1}{x}\right)^{B} = \frac{(1+x)^{A+B}}{x^B}.$$

We shall give here the means, variances and covariances obtained from (3.4)

$$E(r_{1i}) = (n_2+1)^{(2)} n_1^{(i)} / n^{(i+1)}, \tag{3.6}$$

$$\sigma_{ij} = \frac{n_2^{(2)} (n_2+1)^{(2)} n_1^{(i+j)}}{n^{(i+j+2)}} - \frac{n_2^2 (n_2+1)^2 n_1^{(i)} n_1^{(j)}}{n^{(i+1)} n^{(j+1)}}, \tag{3.7}$$

$$\sigma_{ii} = \frac{n_2^{(2)} (n_2+1)^{(2)} n_1^{(2i)}}{n^{(2i+2)}} + \frac{(n_2+1)^{(2)} n_1^{(i)}}{n^{(i+1)}} \left(1 - \frac{(n_2+1)^{(2)} n_1^{(i)}}{n^{(i+1)}}\right). \tag{3.8}$$

These will be needed in the section dealing with asymptotic distributions. The moments for the distribution (2.6) follow at once from (3.3) as

$$E\left(\prod_{ij} r_{1i}^{(a_i)} r_{2j}^{(b_j)}\right) = \sum_{r_1, r_2} r_1^{(\Sigma a_i)} r_2^{(\Sigma b_j)} \tag{3.9}$$
$$\cdot \binom{n_1 - \sum i a_i - 1}{r_1 - \sum a_i - 1} \binom{n_2 - \sum j b_j - 1}{r_2 - \sum b_j - 1} F(r_1, r_2) \Big/ \binom{n}{n_1}.$$

The summation on r_2 is accomplished by putting $r_2 = r_1 - 1, r_1$, and $r_1 + 1$, but after that has been done it is necessary to expand the product of the two factorial factors in factorial powers of the lower index of one of the binomial coefficients. This is easily done for the first few moments, but there appears to be no simple expression for the general case. The means, variances and covariances of r_{1i} are given by (3.6), (3.7) and (3.8) and those of r_{2i} are obtained from these equations by interchanging n_1 and n_2. The other covariances are

$$\sigma_{r_{1i} r_{2j}} = \frac{n_1^{(i+2)} n_2^{(j+2)}}{n^{(i+j+2)}} + 4 \frac{n_1^{(i+1)} n_2^{(j+1)}}{n^{(i+j+1)}} \tag{3.10}$$
$$+ 2 \frac{n_1^{(i)} n_2^{(j)}}{n^{(i+j)}} - \frac{(n_1 + 1)^{(2)} (n_2 + 1)^{(2)} n_1^{(i)} n_2^{(j)}}{n^{(i+1)} n^{(j+1)}}.$$

A slight variation of the method above will give the moments of the s_{1i} in the distribution (2.13). An accent on a summation sign will indicate that the term corresponding to $i = k$ is to be omitted. Differentiating

$$\varphi(t_i) = [t_1 x + t_2 x^2 + \cdots + t_{k-1} x^{k-1} + t_k (x^k + x^{k+1} + \cdots)]^{s_1}$$

a_i times with respect to t_i and finding the coefficient of x^{n_1} after putting $t_i = 1$, we obtain

$$\sum_{\Sigma' i s_{1i} = A} \prod_1^k s_{1i}^{(a_i)} \begin{bmatrix} s_1 \\ s_{1i} \end{bmatrix} \binom{n_1 - A - (k-1) s_{1k} - 1}{s_{1k} - 1} \tag{3.11}$$
$$= s_1^{(\Sigma a_i)} \binom{n_1 - \sum i a_i + a_k - 1}{s_1 - \sum' a_i - 1}.$$

This with (2.13) gives by the same steps as used in obtaining (3.4)

$$E\left(\prod_1^k s_{1i}^{(a_i)}\right) = (n_2 + 1)^{(\Sigma a_i)} \binom{n - \sum i a_i - \sum' a_i}{n_1 - \sum i a_i} \Big/ \binom{n}{n_1}. \tag{3.12}$$

The first two moments are

$$E(s_{1k}) = \frac{(n_2 + 1) n_1^{(k)}}{n^{(k)}}, \tag{3.13}$$

$$\sigma_{ik} = \frac{n_2^2 (n_2 + 1) n_1^{(i+k)}}{n^{(i+k+1)}} - \frac{n_2 (n_2 + 1)^2 n_1^{(i)} n_1^{(k)}}{n^{(i+1)} n^{(k)}}, \tag{3.14}$$

$$\sigma_{kk} = \frac{(n_2+1)^{(2)} n_1^{(2k)}}{n^{(2k)}} + \frac{(n_2+1)n_1^{(k)}}{n^{(k)}}\left(1 - \frac{(n_2+1)n_1^{(k)}}{n^{(k)}}\right). \tag{3.15}$$

The others are, of course, given by (3.6), (3.7) and (3.8).

The joint moments of the variables in (2.14) as obtained from (3.11) are

$$\begin{aligned} E\left(\prod_{ij} s_{1i}^{(a_i)} s_{2j}^{(b_j)}\right) = \sum_{s_1, s_2} s_1^{(\Sigma a_i)} s_2^{(\Sigma b_j)} &\binom{n_1 - \sum i a_i + a_k - 1}{s_1 - \sum' a_i - 1} \\ &\cdot \binom{n_2 - \sum j b_j + b_h - 1}{s_2 - \sum' b_j - 1} F(s_1, s_2) \Big/ \binom{n}{n_1}. \end{aligned} \tag{3.16}$$

In addition to the covariances (3.10) we shall need

$$\begin{aligned} \sigma_{s_{1k}s_{2j}} = \frac{n_1^{(k+2)} n_2^{(j+1)} + 2n_1^{(k+1)} n_2^{(j+1)}}{n^{(k+j+1)}} &+ 2\,\frac{n_1^{(k+1)} n_2^{(j)} + n_1^{(k)} n_2^{(j)}}{n^{(k+j)}} \\ &- \frac{(n_1+1)^{(2)}(n_2+1)^{(2)} n_1^{(k)} n_2^{(j)}}{n^{(k)} n^{(j+1)}}, \end{aligned} \tag{3.17}$$

$$\sigma_{s_{1k}s_{2h}} = \frac{n_1^{(k+1)} n_2^{(h+1)}}{n^{(k+h)}} + 2\,\frac{n_1^{(k)} n_2^{(h)}}{n^{(k+h-1)}} - \frac{(n_1+1)(n_2+1)n_1^{(k)} n_2^{(h)}}{n^{(k)} n^{(h)}}. \tag{3.18}$$

The moments of r in the distribution (2.11) may be derived easily by means of (3.5) as

$$E(r_1^{(a)}) = (n_2+1)^{(a)} \binom{n-a}{n_1-a} \Big/ \binom{n}{n_1}. \tag{3.19}$$

From which

$$E(r_1) = \frac{(n_2+1)n_1}{n}, \tag{3.20}$$

$$\sigma_{r_1}^2 = \frac{(n_2+1)^{(2)} n_1^{(2)}}{n n^{(2)}}. \tag{3.21}$$

4. Distribution and moments of runs of k kinds of elements. This section is a generalization of the preceeding two sections to several kinds of elements. The case $k = 2$ was treated separately because the special character of the function $F(r_1, r_2)$ in this instance made the distribution comparatively simple. Now we shall be interested in k kinds of elements denoted by $a_1, \cdots, a_k$ and we shall suppose there are n_i elements of the ith kind. We let r_{ij} denote the number of runs of elements of the ith kind of length j, and put

$$n = \sum_1^k n_i, \qquad r_i = \sum_{j=1}^{n_i} r_{ij}.$$

The same argument as was used in deriving (2.6) gives

$$P(r_{ij}) = \frac{\prod_{i=1}^{k} \begin{bmatrix} r_i \\ r_{ij} \end{bmatrix} F(r_1, r_2, \cdots, r_k)}{\begin{bmatrix} n \\ n_i \end{bmatrix}} \tag{4.1}$$

where the function $F(r_1, r_2, \cdots, r_k)$, which will be referred to hereafter simply as $F(r_i)$, represents the number of different arrangements of r_1 objects of one kind, r_2 objects of a second kind, and so forth, such that no two adjacent objects are of the same kind. We shall be able to give the explicit expression for $F(r_i)$ after examining the marginal distribution $P(r_i)$. This is obtained by summing (4.1) over r_i with r_{ij} fixed by means of the identity (2.7) giving

$$P(r_i) = \frac{\prod_{i=1}^{k} \binom{n_i - 1}{r_i - 1} F(r_i)}{\begin{bmatrix} n \\ n_i \end{bmatrix}}. \tag{4.2}$$

Despite our present meager knowledge of $F(r_i)$ it is possible to find the moments of the r_i as distributed by (4.2). Since $\sum_{r_i} P(r_i) = 1$, we have the identity

$$\sum_{r_i} \Pi \binom{n_i - 1}{r_i - 1} F(r_i) = \begin{bmatrix} n \\ n_i \end{bmatrix}. \tag{4.3}$$

From this the moments are easily derived. If we put

$$u_i = n_i - r_i \tag{4.4}$$

we have

$$\begin{aligned} \sum_{r_i} \prod u_i^{(a_i)} \prod \binom{n_i - 1}{r_i - 1} F(r_i) &= \sum_{r_i} \prod (n_i - r_i)^{(a_i)} \prod \binom{n_i - 1}{r_i - 1} F(r_i) \\ &= \sum_{r_i} \prod (n_i - 1)^{(a_i)} \prod \binom{n_i - a_i - 1}{r_i - 1} F(r_i) \\ &= \prod (n_i - 1)^{(a_i)} \sum_{r_i} \prod \binom{n_i - a_i - 1}{r_i - 1} F(r_i) \\ &= \prod_{i=1}^{k} (n_i - 1)^{(a_i)} \begin{bmatrix} n - \sum a_i \\ n_i - a_i \end{bmatrix}. \end{aligned}$$

The summation involved in the last step is given by (4.3). On dividing the last equation by $\begin{bmatrix} n \\ n_i \end{bmatrix}$ we get the factorial moments of the u_i

$$E\left(\prod_{1}^{k} u_i^{(a_i)}\right) = \prod_{1}^{k} (n_i - 1)^{(a_i)} \begin{bmatrix} n - \Sigma a_i \\ n_i - a_i \end{bmatrix} \Big/ \begin{bmatrix} n \\ n_i \end{bmatrix}. \tag{4.5}$$

From these equations the moments of the r_i may be found; the means, variances and covariances are

$$E(r_i) = \frac{n_i(n - n_i + 1)}{n}, \tag{4.6}$$

$$\sigma_{ij} = \frac{n_i^{(2)} n_j^{(2)}}{nn^{(2)}}, \tag{4.7}$$

$$\sigma_{ii} = \frac{n_i^{(2)}(n - n_i + 1)^{(2)}}{nn^{(2)}}. \tag{4.8}$$

It is clear that

$$\begin{aligned} \varphi(t_i) &= \text{Coefficient of } \prod_1^k x_i^{n_i} \text{ in} \\ &(x_1 + \cdots + x_k)^k \prod_1^k (x_1 + \cdots + x_{i-1} + t_i x_i + x_{i+1} + \cdots + x_k)^{n_i - 1} \Big/ \begin{bmatrix} n \\ n_i \end{bmatrix} \end{aligned} \tag{4.9}$$

is a generating function for the moments of the variables u_i. This generating function will enable us to find the exact expression for $F(r_i)$ for we have

$$\begin{aligned} P(u_i = n_{ii}) &= \text{Coefficient of } \prod_1^k t_i^{n_{ii}} \text{ in } \varphi(t_i) \\ &= \sum_{\substack{\alpha_i, n_{ij} \\ \Sigma_i n_{ij} = n_j - \alpha_j}} \begin{bmatrix} k \\ \alpha_i \end{bmatrix} \prod_{i=1}^k \begin{bmatrix} n_i - 1 \\ n_{ij} \end{bmatrix} \Big/ \begin{bmatrix} n \\ n_i \end{bmatrix}. \end{aligned}$$

Also

$$P(u_i) = \prod_1^k \binom{n_i - 1}{r_i - 1} F(r_i) \Big/ \begin{bmatrix} n \\ n_i \end{bmatrix}$$

and equating the expressions on the right of the last two equations we have

$$F(r_i) = \frac{\displaystyle\sum_{\alpha_i, n_{ij}} \begin{bmatrix} k \\ \alpha_i \end{bmatrix} \prod_{i=1}^k \begin{bmatrix} n_i - 1 \\ n_{ij} \end{bmatrix}}{\displaystyle\prod_1^k \binom{n_i - 1}{r_i - 1}} \tag{4.10}$$

$$= \sum_{\substack{\alpha_i, n_{ij} \\ \Sigma_i n'_{ij} = r_j - \alpha_j}} \begin{bmatrix} k \\ \alpha_i \end{bmatrix} \prod_{i=1}^k \begin{bmatrix} r_i - 1 \\ n'_{ij} \end{bmatrix} \tag{4.11}$$

in which the prime on the n'_{ij} indicates that the indices corresponding to $j = i$ are to be omitted; hence i takes all the values $1, 2, \cdots, k$ and j takes all values $1, 2, \cdots, k$ except i because the index n_{ii} has been cancelled with $n_i - r_i$ in the binomial coefficient in the denominator of (4.10). It is clear from (4.11) that $F(r_i)$ may be expressed as follows

$$\begin{aligned} F(r_i) = CT \prod_1^k x_i^{-r_i}(x_1 + \cdots + x_k)^k (x_2 + x_3 + \cdots + x_k)^{r_1 - 1} \\ (x_1 + x_3 + \cdots + x_k)^{r_2 - 1} \cdots (x_1 + \cdots + x_{k-1})^{r_k - 1} \end{aligned} \tag{4.12}$$

in which "CT" is an abbreviation for "constant term of."

We are now in a position to obtain moments of the variables r_{ij} in the distribution (4.1) by means of identities similiar to (4.3). As an illustration we compute

$$\sum_{r_i}\binom{n_1-a-1}{r_1-a-1}\prod_2^k\binom{n_i-1}{r_i-1}F(r_i) = \sum_{r_i}\binom{n_1-a-1}{r_1-a-1}\prod_2^k\binom{n_i-1}{r_i-1}$$
$$\cdot CT\prod_1^k x_i^{-r_i}\prod_1^k(x_1+\cdots+t_ix_i+\cdots+x_k)^{r_i-1}\Big]_{t_i=0}$$
$$= CT\prod_1^k x_i^{-n_i}(x_1+\cdots+x_k)^{n-a}(x_2+\cdots+x_k)^a$$
$$= \begin{bmatrix} n \\ n_i\end{bmatrix}\frac{(n-n_1)^{(a)}}{n^{(a)}}$$

or

$$\sum_{r_i}\binom{n_1-a-1}{r_1-a-1}\prod_2^k\binom{n_i-1}{r_i-1}F(r_i) = \frac{(n-a)!\,(n-n_1)^{(a)}}{\prod_1^k n_i!}. \tag{4.13}$$

The moments of r_{ij} may be computed from identities of this type together with (3.3). The first two moments are

$$E(r_{ij}) = (n-n_i+1)^{(2)}n_i^{(j)}/n^{(j+i)} \tag{4.14}$$

$$E(r_{ij}^{(2)}) = n_i^{(2j)}(n-n_i)^{(2)}(n-n_i+1)^{(2)}/n^{(2j+2)} \tag{4.15}$$

$$E(r_{ij}r_{it}) = n_i^{(j+t)}(n-n_i)^{(2)}(n-n_i+1)^{(2)}/n^{(j+t+2)} \qquad j\neq t \tag{4.16}$$

$$\begin{aligned}
E(r_{ij}r_{st}) &= (n_i-j-1)(n_s-t-1)\frac{n_i^{(j-1)}n_s^{(t-1)}}{n^{(j+t+2)}}\{(n_i-j+1)^{(2)}(n_s-t+1)^{(2)}\\
&+2(n-n_i-n_s)(n_i-j+1)(n_s-t+1)(n_s-t+n_i-j)\\
&+(n-n_i-n_s)^{(2)}[(n_s-t+1)^{(2)}+2(n_i-j+1)(n_s-t+1)\\
&+(n_i-j+1)^{(2)}]+2(n-n_i-n_s)^{(3)}(n_i-j+n_s-t+2)\\
&+(n-n_i-n_s)^{(4)}\}+2(n_i-j-1)\frac{n_i^{(j-1)}n_s^{(t-1)}}{n^{(j+t+1)}}\{(n_i-j+1)\\
&\cdot(n_s-t+1)^{(2)}+(n-n_i-n_s)[2(n_i-j+1)(n_s-t+1)\\
&+(n_s-t+1)^{(2)}]+(n-n_i-n_s)^{(2)}[2(n_s-t+1)+(n_i-j+1)]\\
&+(n-n_i-n_s)^{(3)}\}+2(n_s-t-1)\frac{n_i^{(j-1)}n_s^{(t-1)}}{n^{(j+t+1)}}\{(n_s-t+1)\\
&\cdot(n_i-j+1)^{(2)}+(n-n_i-n_s)[2(n_i-j+1)(n_s-t+1)\\
&+(n_i-j+1)^{(2)}]+(n-n_i-n_s)^{(2)}[2(n_i-j+1)+(n_s-t+1)]\\
&+(n-n_i-n_s)^{(3)}\}+4\frac{n_i^{(j-1)}n_s^{(t-1)}}{n^{(j+t)}}\{(n_i-j+1)(n_s-t+1)\\
&+(n-n_i-n_s)(n_i-j+n_s-t+2)+(n-n_i-n_s)^{(2)}\}.
\end{aligned} \tag{4.17}$$

Such a lengthy expression as this last one can hardly be useful to the statistician, and for this reason we shall not define variables s_{ij} analogous to the s_{1i} and s_{2j} of Section 2 and take the time and space to find their moments.

5. Asymptotic distributions. We shall show that some of the distributions obtained previously are asymptotically normal when the n_i become large in such a way that the ratios n_i/n remain fixed. The description "asymptotically normal" means that the distribution approaches the normal distribution uniformly over any finite region as $n_i \to \infty$. The ratios n_i/n will be denoted by e_i, hence $\Sigma e_i = 1$. The symbol $O(1/n^a)$ will represent any function such that

$$\lim_{n\to\infty} n^a O\left(\frac{1}{n^a}\right) = L < \infty.$$

We shall not, of course, be able to get any limit theorems for distributions like (2.6) or (2.9) because the number of independent variables increases with n. We shall consider first the distribution (2.13) whose asymptotic character is given in the following theorem.

THEOREM 1. *The variables*

$$(5.1)\qquad \begin{aligned} x_i &= \frac{s_{1i} - ne_1^i e_2^2}{\sqrt{n}} \qquad i < k \\ x_k &= \frac{s_{1k} - ne_1^k e_2}{\sqrt{n}} \end{aligned}$$

are asymptotically normally distributed with zero means and variances and covariances

$$(5.2)\qquad \begin{aligned} \sigma_{ij} &= e_1^{i+j-1}e_2^3[(i+1)(j+1)e_1e_2 - ije_2 - 2e_1],\ i, j < k,\ i \neq j \\ \sigma_{ii} &= e_1^{2i-1}e_2^3[(i+1)^2e_1e_2 - i^2e_2 - 2e_1] + e_1^i e_2^2,\ i < k \\ \sigma_{ik} &= e_1^{i+k-1}e_2^2[(i+1)ke_1e_2 - ike_2 - e_1],\ i < k \\ \sigma_{kk} &= e_1^{2k-1}e_2[k^2(e_1 - 1)e_2 - e_1] + e_1^k e_2 . \end{aligned}$$

The limiting means, variances and covariances are obtained from the relations (3.6), (3.7), (3.8), (3.13), (3.14) and (3.15).

To demonstrate this theorem we make the substitutions

$$(5.3)\qquad \begin{aligned} n_i &= ne_i & i &= 1, 2 \\ s_{1i} &= ne_1^i e_2^2 + \sqrt{n}x_i & i &= 1, 2, \cdots, k-1 \\ s_{1k} &= ne_1^k e_2 + \sqrt{n}\, x_k \\ s_1 &= ne_1e_2 + \sqrt{n} \sum_1^k x_i \\ A &= n(e_1 - e_1^k - ke_1^k e_2) + \sqrt{n} \sum_1^{k-1} ix_i \end{aligned}$$

in (2.13), and estimate the factorials by means of Stirling's formula

$$m! = \sqrt{2\pi}m^{m+\frac{1}{2}}e^{-m}\left(1 + O\left(\frac{1}{m}\right)\right). \tag{5.4}$$

The result is an unwieldy expression which we shall not present at the moment. First we note that the exponential factors cancel out because the sum of the lower indices of a binomial or multinomial coefficient is equal to the upper index. Also we simplify the expression by considering in detail only terms which involve the x_i; the normalizing constant can be determined from the final limit function. Any function of the parameters will be represented by the letter K. Thus in (5.4) we need consider only the factor $m^{m+\frac{1}{2}}$. All factorials will be of the form

$$m = na + \sqrt{n}L(x) + b \tag{5.5}$$

where $L(x)$ is a linear function of the x_i, and a and b are independent of n and x_i. Now

$$\begin{aligned} m^{m+\frac{1}{2}} &= (na + \sqrt{n}L(x) + b)^{na+\sqrt{n}L(x)+b+\frac{1}{2}} \\ &= (na)^{na+\sqrt{n}L(x)+b+\frac{1}{2}}\left(1 + \frac{L(x)}{a\sqrt{n}} + \frac{b}{an}\right)^{na+\sqrt{n}L(x)+b+\frac{1}{2}} \\ &= K(na)^{\sqrt{n}L(x)}\left(1 + \frac{L(x)}{a\sqrt{n}} + \frac{b}{an}\right)^{na+\sqrt{n}L(x)+b+\frac{1}{2}} \end{aligned}$$

and

$$\begin{aligned} \log m^{m+\frac{1}{2}} &= K + \sqrt{n}L(x)\log na + (na + \sqrt{n}L(x) + b + \tfrac{1}{2}) \\ &\qquad \cdot \log\left(1 + \frac{L(x)}{a\sqrt{n}} + \frac{b}{an}\right) \\ &= K + \sqrt{n}L(x)\log na + (na + \sqrt{n}L(x) + b + \tfrac{1}{2}) \\ &\qquad \cdot\left(\frac{L(x)}{a\sqrt{n}} + \frac{b}{an} - \frac{L^2(x)}{a^2 n} + O\left(\frac{1}{n^{3/2}}\right)\right) \\ &= K + \sqrt{n}L(x)(1 + \log na) + \frac{1}{2a}L^2(x) + O\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{5.6}$$

so terms arising from b (and $b + \frac{1}{2}$ in the exponent) will be neglected as they give rise only to terms independent of the x_i or of order $1/n^{\frac{1}{2}}$. Of course $\log (1 + O(1/m)) = O(1/m)$. Thus, keeping significant terms only, the result of the substitutions (5.3) and (5.4) in (2.13) after taking logarithms and using (5.6) is

$$\begin{aligned} -\log P(r_i) &= K + \sqrt{n}\sum_1^{k-1} x_i(\log ne_1^i e_2^2 + 1) + \sum_1^{k-1}\frac{x_i^2}{2e_1^i e_2^2} \\ &\quad - \sqrt{n}\left(\sum_1^k x_i\right)(\log ne_2^2 + 1) + \frac{1}{2e_2^2}\left(\sum_1^k x_i\right)^2 \end{aligned}$$

$$+\sqrt{n}\left(\sum_{1}^{k-1} ix_i + (k-1)x_k\right)(\log ne_1^k + 1) - \frac{1}{2e_1^k}\left(\sum_{1}^{k} ix_i + (k-1)x_k\right)^2 \tag{5.7}$$
$$+ 2\sqrt{n}x_k(\log ne_1^k e_2 + 1) + \frac{x_k^2}{e_1^k e_2} - \sqrt{n}\left(\sum_{1}^{k} ix_i\right)(\log ne_1^{k+1} + 1)$$
$$+ \frac{1}{2e_1^{k+1}}\left(\sum_{1}^{k} ix_i\right)^2 + O\left(\frac{1}{\sqrt{n}}\right).$$

The coefficients of $x_i (i < k)$ and x_k are

$$\sqrt{n}(\log ne_1^i e_2^2 + 1 - \log ne_2^2 - 1 + i \log ne_1^k + i - i \log ne_1^{k+1} - i) = 0,$$
$$\sqrt{n}(-\log ne_2^2 - 1 + k \log ne_1^k + k + 2 \log ne_1^k e_2 + 2 - k \log ne_1^{k+1} - k) = 0.$$

Hence only the quadratic terms remain and we have

$$-\log P = K + \tfrac{1}{2}\sum_{i,j} \sigma^{ij} x_i x_j + O\left(\frac{1}{\sqrt{n}}\right) \tag{5.8}$$

where

$$\begin{aligned}
\sigma^{ij} &= \frac{1}{e_2^2} + \frac{ije_2}{e_1^{k+1}} && i, j < k,\ i \neq j,\\
\sigma^{ii} &= \frac{1}{e_2^2} + \frac{1}{e_1^i e_2^2} + \frac{i^2 e_2}{e_1^{k+1}} && i < k,\\
\sigma^{ik} &= \frac{1}{e_2^2} + \frac{i + i(k-1)e_2}{e_1^{k+1}} && i < k,\\
\sigma^{kk} &= \frac{1}{e_2^2} + \frac{2}{e_1^k e_2} + \frac{k^2}{e_1^{k+1}} - \frac{(k-1)^2}{e_1^k}.
\end{aligned} \tag{5.9}$$

It is merely a matter of straightforward multiplication of the two matrices to verify that $\| \sigma^{ij} \|$ is the inverse of $\| \sigma_{ij} \|$, hence is a positive definite matrix. The details of the verification will be omitted. We have then

$$P = Ke^{-\frac{1}{2}\Sigma\sigma^{ij}x_ix_j}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right). \tag{5.10}$$

In this equation K must necessarily contain the factor $\left(\frac{1}{\sqrt{n}}\right)^k$ because there are $k + 5$ factorials in the denominator and 5 in the numerator of (2.13). Since $\Delta r_i = 1$, this factor, in view of (5.1), may be replaced by $\Pi\Delta x_i$, so

$$P = Ke^{-\frac{1}{2}\Sigma\sigma^{ij}x_ix_j}\Pi\Delta x_i\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right). \tag{5.11}$$

If we restrict the x_i to any finite region R in the x-space, the function $O(1/n^{\frac{1}{2}})$ approaches zero uniformly as $n \to \infty$. Thus, if $A_i < B_i$ are any positive

numbers such that the corresponding values of x_i, say a_i and b_i, obtained by substituting A_i and B_i for r_i in (5.1), determine a rectangular region $R'(a_i < x_i < b_i)$, which lies in R we have

$$\sum_{r_i=A_i}^{B_i} P(r_i) = \sum_{x_i=a_i}^{b_i} Ke^{-\frac{1}{2}\Sigma\sigma^{ij}x_ix_j}\Pi\Delta x_i\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) \tag{5.12}$$

$$\xrightarrow[n\to\infty]{} \int_{R'} Ke^{-\frac{1}{2}\Sigma\sigma^{ij}x_ix_j}\Pi\, dx_i$$

by the definition of a definite integral and Riemann's fundamental theorem.

We have given some details of this proof in order that it may serve as a model for other theorems of a similiar nature which will appear later, and for which a complete proof will not be given. Two immediate consequences of Theorem 1 will now be stated as corollaries.

COROLLARY 1. *The variable*

$$x = \frac{r - ne_1e_2}{\sqrt{ne_1e_2}}$$

where r is the total number of runs of one kind of element, is asymptotically normally distributed with zero mean and unit variance. The limiting mean and variance were computed from (3.20) and (3.21).

COROLLARY 2. *The variable $Q = \Sigma\sigma^{ij}x_ix_j$ is asymptotically distributed according to the χ^2-law with k degrees of freedom.*

In exactly the same manner in which Theorem 1 was deduced from (2.13), we may prove the following theorem corresponding to the distribution (2.14).

THEOREM 2. *The variables*

$$\begin{aligned} x_i &= \frac{s_{1i} - ne_1^ie_2^2}{\sqrt{n}} & i < k, \\ x_k &= \frac{s_{1k} - ne_1^ke_2}{\sqrt{n}}, & \\ y_i &= \frac{s_{2i} - ne_1^2e_2^i}{\sqrt{n}} & i < h, \end{aligned} \tag{5.13}$$

are asymptotically normally distributed with zero means and variances and covariances

$$\begin{aligned} \sigma_{x_ix_j} &= e_1^{i+j-1}e_2^3[(i + 1)(j + 1)e_1e_2 - ije_2 - 2e_1] & i, j < k, \\ \sigma_{x_ix_i} &= e_1^{2i-1}e_2^3[(i + 1)^2e_1e_2 - i^2e_2 - 2e_1] + e_1^ie_2^2 & i < k, \\ \sigma_{x_ix_k} &= e_1^{i+k-1}e_2^2[(i + 1)ke_1e_2 - ike_2 - e_1] & i < k, \\ \sigma_{x_kx_k} &= e_1^{2k-1}e_2[- k^2e_2^2 - e_1] + e_1^ke_2, & \end{aligned}$$

$$
\begin{aligned}
(5.14)\quad \sigma_{y_iy_j} &= e_2^{i+j-1}e_1^3[(i+1)(j+1)e_1e_2 - ije_1 - 2e_2] && i, j < h,\\
\sigma_{y_iy_i} &= e_2^{2i-1}e_1^3[(i+1)^2e_1e_2 - i^2e_1 - 2e_2] + e_2^ie_1^2 && i < h,\\
\sigma_{x_iy_j} &= e_1^{i+1}e_2^{j+1}[(i+1)(j+1)e_1e_2 - 2ie_2 - 2je_1 + 4e_1e_2 + 2] && \\
& && i < k, j < h,\\
\sigma_{x_ky_j} &= e_1^{k+1}e_2^j[k(j+1)e_1e_2 - 2(k-1)e_2 - (j-1)e_1 + 2e_1e_2] && j < h.
\end{aligned}
$$

These limiting variances were computed from the variances and covariances given in Section 3. We have chosen the variable s_{2h} of (2.14) as the dependent variable. The proof of this theorem is omitted. From it the following corollaries are deduced immediately.

COROLLARY 3. *If $u_i = x_i$ and $u_{k+i} = y_i$ of (5.13) and $\| \sigma^{ij} \|$ $(i, j = 1, 2, \cdots, k + h - 1)$ denotes the inverse of (5.14), then the variable $Q = \Sigma\sigma^{ij}u_iu_j$ is asymptotically distributed according to the χ^2-law with $k + h - 1$ degrees of freedom.*

COROLLARY 4. *If $s_i = s_{1i} + s_{2i}$ denotes the total number of runs of both kinds of elements of length i, and s_k the total number of runs of length greater than $k - 1$, then the variables*

$$
(5.15)\qquad
\begin{aligned}
x_i &= \frac{s_i - n(e_1^ie_2^2 + e_2^ie_1^2)}{\sqrt{n}} \qquad i < k\\
x_k &= \frac{s_k - n(e_1^ke_2 + e_2^ke_1)}{\sqrt{n}}
\end{aligned}
$$

are asymptotically normally distributed with zero means and variances and covariances

$$
(5.16)\qquad \sigma_{ij} = \sigma_{x_ix_j} + \sigma_{x_iy_j} + \sigma_{x_jy_i} + \sigma_{y_iy_j}\,.
$$

We have put $h = k$ in Theorem 2 to obtain this result. The terms on the right of (5.16) are defined by (5.14); terms which do not appear there may be found by interchanging e_1 and e_2 in one of the relations. For example $\sigma_{y_ky_k}$ is given by interchanging e_1 and e_2 in the fourth equation of the set (5.14).

COROLLARY 5. *The variable $Q = \Sigma\sigma^{ij}x_ix_j$ where the x_i are defined by (5.15) and $\| \sigma^{ij} \|$ is the inverse of (5.16), is asymptotically distributed according to the χ^2-law with k degrees of freedom.*

COROLLARY 6. *If s denotes the total number of runs of both kinds of elements, then the variable*

$$
x = \frac{s - 2ne_1e_2}{2\sqrt{n}e_1e_2}
$$

is asymptotically normally distributed with zero mean and unit variance. This is the result derived by Wald and Wolfowitz [13].

6. Asymptotic distributions for k kinds of elements. We now investigate the asymptotic character of the distribution (4.2)

$$P(r_i) = \frac{\prod_{i=1}^{k} \binom{n_i - 1}{r_i - 1} F(r_i)}{\begin{bmatrix} n \\ n_i \end{bmatrix}} \tag{6.1}$$

where r_i is the total number of runs of the ith kind of element.

THEOREM 1. *If $k > 2$, the variables*

$$x_i = \frac{r_i - ne_i(1 - e_i)}{\sqrt{n}} \tag{6.2}$$

are asymptotically normally distributed with zero means and variances and covariances

$$\sigma_{ij} = e_i^2 e_j^2, \qquad \sigma_{ii} = e_i^2(1 - e_i)^2. \tag{6.3}$$

The restriction $k > 2$ is made because in the case $k = 2$ the correlation between the two variables approaches one, and the numbers σ_{ij} are all equal. The result may be called a degenerate normal distribution and might be included in the theorem in this sense; we have chosen to omit it because this case is better taken care of by Corollary 1 of the previous section.

The proof of this theorem will be simplified if in the moments (4.5) we replace the numbers $n_i - 1$ by n_i. This substitution will not, of course, affect the limiting moments. Hence we consider the variables v_i with moments given by

$$E\left(\prod_1^k v_i^{(a_i)}\right) = \frac{\prod_1^k n_i^{(a_i)} \begin{bmatrix} n - \Sigma a_i \\ n_i - a_i \end{bmatrix}}{\begin{bmatrix} n \\ n_i \end{bmatrix}} \tag{6.4}$$

and shall show that

$$y_i = \frac{v_i - ne_i^2}{\sqrt{n}} \tag{6.5}$$

are asymptotically normally distributed with zero means and variances and covariance (6.3). It is possible to prove this statement by showing that the characteristic function (Fourier transform) obtained by substituting $i\theta_i$ for t_i in the moment generating function

$$\begin{aligned} \varphi_n(t_i) = \text{Coef. of } \prod_1^k x_i^{n_i} \text{ in} \\ \prod_1^k (x_1 + \cdots + x_{i-1} + t_i x_i + x_{i+1} + \cdots + x_k)^{n_i} \Big/ \begin{bmatrix} n \\ n_i \end{bmatrix} \end{aligned} \tag{6.6}$$

approaches

$$\varphi(\theta_i) = e^{-\frac{1}{2}\Sigma\sigma_{ij}\theta_i\theta_j}$$

as $n \to \infty$. This method is not appropriate for proving a similiar theorem which appears in Part II, and we prefer to give here a demonstration that will suffice for both theorems.

In order to prove our theorem we consider the general term in the coefficient of $\Pi x_i^{n_i}$ in (6.6)

$$C(m_{ij}) = \prod_{i=1}^{k} \begin{bmatrix} n_i \\ m_{ij} \end{bmatrix} \Pi\, t_i^{m_{ii}} \Big/ \begin{bmatrix} n \\ n_i \end{bmatrix} \tag{6.7}$$

in which

$$\sum_{i=1}^{k} m_{ij} = n_j \tag{6.8}$$

must be required as well as the usual restriction on indices of a multinomial coefficient, $\sum_{j=1}^{k} m_{ij} = n_i$. Therefore only $(k-1)^2$ of the indices are independent. Clearly $m_{ii} = v_i$. Now without concerning ourselves about the statistical significance of the variables m_{ij}, let us consider their distribution

$$D(m_{ij}) = \prod_{i=1}^{k} \begin{bmatrix} n_i \\ m_{ij} \end{bmatrix} \Big/ \begin{bmatrix} n \\ n_i \end{bmatrix} \tag{6.9}$$

in which the variables corresponding to the values $i, j = 1, 2, \cdots, k-1$ will be chosen as the independent ones. We shall now prove a theorem from which Theorem 1 follows immediately.

THEOREM 2. *The variables*

$$x_{ij} = \frac{m_{ij} - ne_ie_j}{\sqrt{n}} \qquad i, j = 1, 2, \cdots, k-1 \tag{6.10}$$

are asymptotically normally distributed with zero means and variances and co-variances given by

$$\begin{aligned} \sigma_{ij,pq} &= e_ie_je_pe_q\,, \\ \sigma_{ij,ip} &= -\,e_i(1-e_i)e_je_p\,, \\ \sigma_{ij,ij} &= e_ie_j(1-e_i)(1-e_j). \end{aligned} \tag{6.11}$$

First it is to be noted that the moments of the m_{ij} are easily obtained from the identity

$$\sum_{\Sigma_i m_{ij} = n_j} \prod_{i=1}^{k} \begin{bmatrix} n_i \\ m_{ij} \end{bmatrix} = \begin{bmatrix} n \\ n_i \end{bmatrix} \tag{6.12}$$

as follows

$$\begin{aligned} \sum \prod_{ij} m_{ij}^{(a_{ij})} \prod_i \begin{bmatrix} n_i \\ m_{ij} \end{bmatrix} &= \sum \prod_i n_i^{(\Sigma_j a_{ij})} \prod_i \begin{bmatrix} n_i - \sum_j a_{ij} \\ m_{ij} - a_{ij} \end{bmatrix} \\ &= \prod_i n_i^{(\Sigma_j a_{ij})} \begin{bmatrix} n - \sum_{ij} a_{ij} \\ n_j - \sum_i a_{ij} \end{bmatrix} \end{aligned}$$

and on dividing this last relation by $\begin{bmatrix} n \\ n_i \end{bmatrix}$ we obtain

$$E(\prod_{i,j} m_{ij}^{(a_{ij})}) = \prod_i n_i^{(\Sigma_j a_{ij})} \prod_j n_i^{(\Sigma_i a_{ij})} / n^{(\Sigma_{ij} a_{ij})} \tag{6.13}$$

from which the moments (6.11) and the means in (6.10) were computed.

The proof of the theorem is similar to that of Theorem 1 in Section 5. We make the substitutions

$$n_i = ne_i, \qquad m_{kj} = n_j - \sum_{i=1}^{k-1} m_{ij},$$

$$m_{ik} = n_i - \sum_{j=1}^{k-1} m_{ij}, \qquad m_{kk} = 2n_k + \sum_{i,j=1}^{k-1} m_{ij} - n,$$

$$m_{ij} = ne_ie_j + \sqrt{n}x_{ij},$$

in (6.9) and employ Stirling's formula exactly as before. The details are too similiar to warrant repetition. The final result is

$$D(m_{ij}) = Ke^{-\frac{1}{2}\Sigma\sigma^{ij,pq}x_{ij}x_{pq}}\Pi\, dx_{ij}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right). \tag{6.14}$$

Where $\| \sigma^{ij,pq} \|$ is the inverse of (6.11) and is defined by

$$\sigma^{ij,pq} = \frac{1}{e_k^2}, \qquad \sigma^{ij,ij} = \frac{1}{e_k^2} + \frac{1}{e_ie_k} + \frac{1}{e_je_k} + \frac{1}{e_ie_j},$$

$$\sigma^{ij,ip} = \frac{1}{e_1e_k} + \frac{1}{e_k^2}, \qquad \sigma^{ij,pj} = \frac{1}{e_1e_k} + \frac{1}{e_k^2}.$$

Theorem 1 is a corollary of Theorem 2. Also we may state these additional results:

Corollary 1. *If k (≥ 3) kinds of elements are arranged at random and r denotes the total number of runs of all kinds of elements, then the variable*

$$x = \frac{r - n(1 - \Sigma e_i^2)}{\sqrt{n}}$$

is asymptotically normally distributed with zero mean and variance

$$\sigma^2 = \Sigma e_i^2 - 2\Sigma e_i^3 + (\Sigma e_i^2)^2$$

where e_i is the proportion of elements of the i-th kind.

Corollary 2. *The variable $Q = \Sigma\sigma^{ij}x_ix_j$ where the x_i are defined by* (6.2) *and $\| \sigma^{ij} \|$ is the inverse of* (6.3), *is asymptotically distributed according to the χ^2-law with k degrees of freedom.*

As was mentioned in Section 4, we could define variables s_{ij} ($i = 1, 2, \cdots, k$ and $j = 1, 2, \cdots, h_i$, the h_i being a set of k arbitrary integers) with a distribution similiar to (2.14). If one worked through the details he would find, no

doubt, that these variables are asymptotically normal. The matrix of variances and covariances is so complicated, however, that such a theorem would hardly be useful to the statistician, and the author does not feel that it would be worthwhile to go through the long and tedious details merely for the sake of completeness.

Part II

Instead of having the number of elements of each kind fixed, we now suppose that they are randomly drawn from a binomial or multinomial population. The numbers n_i thus become random variables subject only to the restriction that $\Sigma n_i = n$, the sample number. The development will be entirely analogous to that of Part I, and the same notation will be used. The probability associated with the ith kind of element will be denoted by p_i.

7. Distributions and moments. The major part of the derivation of the various distribution functions has already been done in Sections 2 and 3. With the distributions of these sections we need only employ the fundamental relation

$$P(X, Y) = P_1(X \mid Y)P_2(Y) \tag{7.1}$$

in order to obtain the distributions required here. X will represent the set of variables r_{ij} or r_i, and Y the variables n_i. For the binomial population $P_2(Y)$ will be

$$P(n_1, n_2) = \binom{n}{n_1} p_1^{n_1} p_2^{n_2}. \tag{7.2}$$

Therefore we may write down at once the distributions

$$P(r_{ij}, n_i) = \begin{bmatrix} r_1 \\ r_{1j} \end{bmatrix}\begin{bmatrix} r_2 \\ r_{2j} \end{bmatrix} F(r_1, r_2) p_1^{n_1} p_2^{n_2}, \tag{7.3}$$

$$P(r_{1i}, n_i) = \begin{bmatrix} r_1 \\ r_{1j} \end{bmatrix}\binom{n_2 + 1}{r_1} p_1^{n_1} p_2^{n_2}, \tag{7.4}$$

$$P(r_1, n_i) = \binom{n_1 - 1}{r_1 - 1}\binom{n_2 + 1}{r_1} p_1^{n_1} p_2^{n_2}, \tag{7.5}$$

$$P(s_{1j}, n_i) = \begin{bmatrix} s_1 \\ s_{1j} \end{bmatrix}\binom{n_1 - A - (k-1)s_{1k} - 1}{s_{1k} - 1}\binom{n_2 + 1}{s_1} p_1^{n_1} p_2^{n_2}, \tag{7.6}$$

$$\begin{aligned} P(s_{1i}, s_{2j}, n_i) = \begin{bmatrix} s_1 \\ s_{1i} \end{bmatrix}\begin{bmatrix} s_2 \\ s_{2j} \end{bmatrix}&\binom{n_1 - A - (k-1)s_{1k} - 1}{s_{1k} - 1} \\ &\cdot\binom{n_2 - B - (h-1)s_{2h} - 1}{s_{2h} - 1} F(s_1, s_2) p_1^{n_1} p_2^{n_2}, \\ &\qquad i = 1, \cdots, k, j = 1, \cdots, h, \end{aligned} \tag{7.7}$$

corresponding to the distributions (2.6), (2.9), (2.11), (2.13) and (2.14) respectively. Of course there is some dependence among the arguments. In (7.4), for example, n_1 is determined by $\Sigma i r_{1i} = n_1$, and n_2 by $n - n_1 = n_2$. In the last three distributions one of the n_i is independent and one may sum these with respect to n_1 from zero to n and obtain the distributions of the r's alone. The results of such summations are quite cumbersome and in some cases can only be indicated, so we shall retain the n_i as relevant variables. This remark applies also to the multinomial distribution.

We shall obtain expressions for the joint moments of the variables in these distributions. It is clear that the moments in Section 3 will be of considerable aid; for, using the notation of (7.1), we have

$$E(f(X)g(Y)) = \sum_{XY} f(X)g(Y)P(X, Y) = \sum_{Y} g(Y)P_2(Y)[\sum_{X} f(X)P_1(X/Y)] \tag{7.8}$$

and the sum in the bracket on the right has been computed in Section 3. It remains only for us to multiply the previous moments by $g(Y)P_2(Y)$ and sum on Y. Corresponding to (3.4), (3.12), (3.9) and (3.19) we have

$$E\left(n_1^{(a)} \prod_1^{n_1} r_{1i}^{(a_i)}\right) = \sum_{n_1=0}^{n} n_1^{(a)}(n_2+1)^{(\Sigma a_i)} \binom{n - \Sigma i a_i - \Sigma a_i}{n_1 - \Sigma i a_i} p_1^{n_1} p_2^{n_2}, \tag{7.9}$$

$$E\left(n_1^{(a)} \prod_1^{k} s_{1i}^{(a_i)}\right) = \sum_{n_1=0}^{n} n_1^{(a)}(n_2+1)^{(\Sigma a_i)} \binom{n - \Sigma i a_i - \Sigma' a_i}{n_1 - \Sigma i a_i} p_1^{n_1} p_2^{n_2}, \tag{7.10}$$

$$E(n_1^{(a)} r_1^{(b)}) = \sum_{n_1=0}^{n} n_1^{(a)}(n_2+1)^{(b)} \binom{n-b}{n_1-b} p_1^{n_1} p_2^{n_2}, \tag{7.11}$$

$$\begin{aligned} E\left(n_1^{(a)} \prod_1^{k} s_{1i}^{(a_i)} \prod_1^{h} s_{2j}^{(b_j)}\right) &= \sum_{n_1, s_1, s_2} n_1^{(a)} s_1^{(\Sigma a_i)} s_2^{(\Sigma b_j)} \binom{n_1 - \Sigma i a_i + a_k - 1}{s_1 - \Sigma' a_i - 1} \\ &\quad \cdot \binom{n_2 - \Sigma j b_j + b_h - 1}{s_2 - \Sigma' b_j - 1} F(s_1, s_2) p_1^{n_1} p_1^{n_2}, \end{aligned} \tag{7.12}$$

for moments from (7.4), (7.6), (7.5) and (7.7) respectively. In order to perform the summations indicated in these last relations it is necessary to expand the factors multiplying the binomial coefficient in factorial powers of its lower index. That is, we must write

$$n_1^{(a)}(n_2+1)^{(b)} = \sum_{i=0}^{a+b} C_i(n, a, b)(n_1 - b)^{(i)}. \tag{7.13}$$

Again it is not possible to give a simple expression for the coefficients $C_i(n, a, b)$ in general, but for the first few moments they present no difficulty. For example from (7.9)

$$
\begin{aligned}
E(n_1 r_{1i}) &= \sum_{n_1=0}^{n} n_1(n - n_1 + 1)\binom{n-i-1}{n_1-i} p_1^{n_1} p_2^{n_2} \\
&= \sum_{n_1} [i(n-i+1) + (n-2i)(n_1-i) + (n_1-i)^{(2)}] \\
&\qquad \cdot \binom{n-i-1}{n_1-i} p_1^{n_1} p_2^{n_2} \\
&= \sum_{n_1} \Bigg[i(n-i+1)\binom{n-i-1}{n_1-i} + (n-2i)(n-i-1) \\
&\qquad \cdot \binom{n-i-2}{n_1-i-1} - (n-i-1)^{(2)}\binom{n-i-3}{n_1-i-2} \Bigg] p_1^{n_1} p_2^{n_2} \\
&= [i(n-i+1) + (n-2i)(n-i-1)p_1 - (n-i-1)^{(2)} p_1^2] p_1^i p_2.
\end{aligned} \tag{7.14}
$$

We give below some means, variances and covariances which will be required later.

$$
\begin{aligned}
E(r_{1i}) &= p_1^i p_2[(n-i-1)p_2 + 2], \\
E(s_{1k}) &= p_1^k[(n-k)p_2 + 1], \\
\sigma_{r_{1i}r_{1j}} &= p_1^{i+j} p_2^2\{(n-i-j)^{(2)} p_2^2 + (n-i-j)p_2(1+5p_1) + 6p_1^2 \\
&\qquad - [(n-i-1)p_2 + 2][(n-j-1)p_2 + 2]\}, \\
\sigma_{r_{1i}r_{1i}} &= p_1^{2i} p_2^2\{(n-2i)^{(2)} p_2^2 + (n-2i)p_2(1+5p_1) + 6p_1^2 \\
&\qquad - [(n-i-1)p_2 + 2]^2\} + p_1^i p_2[(n-i-1)p_2 + 2], \\
\sigma_{r_{1i}r_{2j}} &= p_1^i p_2^j\{(n-i-j-2)^{(2)} p_1^2 p_2^2 + 4(n-i-j-1)p_1 p_2 + 2 \\
&\qquad - [(n-i-1)p_2 + 2][(n-j-1)p_1 + 2], \\
\sigma_{s_{1i}s_{1k}} &= p_1^{i+k} p_2\{(n-i-k+1)^{(2)} - 2(n-i-k)^{(2)} p_1 \\
&\qquad + (n-i-k-1)^{(2)} p_1^2 - [(n-i-1)p_2 + 2][(n-k)p_2 + 1]\}, \\
\sigma_{s_{1k}s_{1k}} &= p_1^{2k}\{(n-2k+1)^{(2)} - 2(n-2k)^{(2)} p_1 + (n-2k)^{(2)} p_1^2 \\
&\qquad - [(n-k)p_2 + 1]^2\} + p_1^k[(n-k)p_2 + 1], \\
\sigma_{s_{1k}s_{2j}} &= p_1^k p_2^j\{(n-k-j-2)^{(2)} p_1^2 p_2 + 2(n-k-j-1)p_1(1+p_2) \\
&\qquad + 2(1+p_1) - p_1[(n-k)p_2 + 1][(n-j-1)p_1 + 2]\}.
\end{aligned} \tag{7.15}
$$

In order to obtain the distribution of runs in samples from a multinomial population, we multiply the distributions of Section 4 by

$$
P(n_i) = \begin{bmatrix} n \\ n_i \end{bmatrix} \prod_1^k p_i^{n_i}. \tag{7.16}
$$

Corresponding to (4.1) and (4.2) then, we have

$$
P(r_{ij}, n_i) = \prod_{i=1}^{k} \begin{bmatrix} r_i \\ r_{ij} \end{bmatrix} F(r_i) \prod_1^k p_i^{n_i} \tag{7.17}
$$

$$
P(r_i, n_i) = \prod_1^k \binom{n_i - 1}{r_i - 1} F(r_i) \prod_1^k p_i^{n_i}. \tag{7.18}
$$

In (7.17) r_{ij} is the number of runs of length j of elements with probability p_i. In (7.18) r_i is the total number of runs of elements with probability p_i. As before, we shall investigate in detail only the distribution (7.18). The moments of n_i and r_i follow at once from (7.8) and (4.5)

$$E\left(\prod_1^k (n_i^{(a_i)} u_i^{(b_i)})\right) = \sum_{n_i} \prod_1^k (n_i^{(a_i)}(n_i - 1)^{(b_i)}) \begin{bmatrix} n - \Sigma b_i \\ n_i - b_i \end{bmatrix} \prod_1^k p_i^{n_i} \tag{7.19}$$

where $u_i = n_i - r_i$. The means, variances and covariances of the r_i are

$$\begin{aligned} E(r_i) &= np_i(1 - p_i) + p_i^2, \\ \sigma_{r_i r_j} &= -np_ip_j(1 - 2p_i - 2p_j + 3p_ip_j) - p_ip_j(2p_i + 2p_j - 5p_ip_j), \\ \sigma_{r_i r_i} &= np_i(1 - 4p_i + 6p_i^2 - 3p_i^3) + p_i^2(3 - 8p_i + 5p_i^2). \end{aligned} \tag{7.20}$$

8. **Asymptotic distributions from binomial population.** We turn our attention first to the distribution (7.7) and state a theorem analogous to Theorem 2 of Section 5.

THEOREM 1. *The variables*

$$\begin{aligned} u_i = x_i &= \frac{s_{1i} - np_1^ip_2^2}{\sqrt{n}}, & i &= 1, \cdots, k-1, \\ u_k = x_k &= \frac{s_{1k} - np_1^kp_2}{\sqrt{n}}, \\ u_{k+i} = y_i &= \frac{s_{2i} - np_1^2p_2^i}{\sqrt{n}}, & i &= 1, \cdots, h-1, \\ u_{k+h} = z &= \frac{n_1 - np_1}{\sqrt{n}}, \end{aligned} \tag{8.1}$$

are asymptotically normally distributed with zero means and variances and covariances

$$\begin{aligned} \sigma_{x_ix_i} &= p_1^ip_2^2 - (2i+1)p_1^{2i}p_2^4 + 2p_1^{2i+1}p_2^3, \\ \sigma_{x_ix_j} &= -(i+j+1)p_1^{i+j}p_2^4 + 2p_1^{i+j+1}p_2^3, \\ \sigma_{x_ix_k} &= -(i+k+1)p_1^{i+k}p_2^3 + p_1^{i+k+1}p_2^2, \\ \sigma_{x_kx_k} &= p_1^kp_2 - (2k+1)p_1^{2k}p_2^2, \\ \sigma_{y_iy_j} &= -(i+j+1)p_1^4p_2^{i+j} + 2p_1^3p_2^{i+j+1}, \\ \sigma_{y_iy_i} &= p_1^2p_2^i - (2i+1)p_1^4p_2^{2i} + 2p_1^3p_2^{2i+1}, \\ \sigma_{x_iy_j} &= -(i+j+3)p_1^{i+2}p_2^{j+2} + 2p_1^{i+1}p_2^{j+1}, \\ \sigma_{x_ky_j} &= -(k+j+2)p_1^{k+2}p_2^{j+1} + p_1^{k+1}p_2^j(1+p_2), \\ \sigma_{x_iz} &= ip_1^ip_2^3 + p_1^{i+1}p_2(1 - 4p_2), \\ \sigma_{x_kz} &= (k+1)p_1^kp_2^2 - p_1^k(1+p_2), \\ \sigma_{y_iz} &= ip_1^3p_2^i + p_1p_2^{i+1}(1 - 4p_1), \\ \sigma_{zz} &= p_1p_2. \end{aligned} \tag{8.2}$$

We have taken s_{2h} and n_2 to be the dependent variables of (7.7). The method of proof of this theorem is the same as that of Theorem 1 in Section 5, and will be omitted. As consequences of the theorem we have

COROLLARY 1. *The variable*

$$Q = \sum_1^{k+h} \sigma^{ij} u_i u_j$$

is asymptotically distributed according to the χ^2-law with $k + h$ degrees of freedom.

COROLLARY 2. *Any subset $u_{i_1}, u_{i_2}, \cdots, u_{i_m}$ of the variables (8.1) is asymptotically normally distributed with zero means and variances and covariances $\| \sigma_{i_j i_k} \|$, and*

$$Q = \sum_{j,k=1}^{m} \sigma^{i_j i_k} u_{i_j} u_{i_k}$$

is asymptotically distributed according to the χ^2-law with m degrees of freedom. $\| \sigma^{i_j i_k} \|$ is the inverse of $\| \sigma_{i_j i_k} \|$.

COROLLARY 3. *If $s_i = s_{1i} + s_{2i}$ represents the total number of runs of length i of both kinds of elements, and s_k the number of runs of length greater than $k - 1$, then the variables*

$$(8.3)\qquad \begin{aligned} x_i &= \frac{s_i - n(p_1^i p_2^2 + p_1^2 p_2^i)}{\sqrt{n}}, \qquad i = 1, \cdots, k-1, \\ x_k &= \frac{s_k - n(p_1^k p_2 + p_1 p_2^k)}{\sqrt{n}}, \end{aligned}$$

are asymptotically normally distributed with zero means and variances and covariances

$$(8.4)\qquad \sigma_{ij} = \sigma_{x_i x_j} + \sigma_{x_i y_j} + \sigma_{x_j y_i} + \sigma_{y_i y_j}$$

where the terms on the right of (8.4) *are defined by* (8.2). We have put $h = k$ in Theorem 1 to obtain this result.

COROLLARY 4. *The variable*

$$(8.5)\qquad Q = \sum_1^k \sigma^{ij} x_i x_j$$

where the x_i are defined by (8.3) *and $\| \sigma^{ij} \|$ is the inverse of* (8.4), *is asymptotically distributed according to the χ^2-law with k degrees of freedom.*

COROLLARY 5. *If r denotes the total number of runs of both kinds of elements, then*

$$(8.6)\qquad x = \frac{r - 2np_1p_2}{2\sqrt{np_1p_2(1 - 3p_1p_2)}}$$

is asymptotically normally distributed with zero mean and unit variance. This is the result obtained by Wishart and Hirshfeld [11].

9. Asymptotic distributions from the multinomial population. In this section we assume $k > 2$ to avoid degenerate distributions. Because of the function $F(r_i)$ in (7.18) we do not investigate this distribution directly, but derive a more general asymptotic distribution as was done in Section 6. We consider the distribution

$$D(m_{ij}, n_i) = \prod_{i=1}^{k} \left(\begin{bmatrix} n_i \\ m_{ij} \end{bmatrix} p_i^{n_i} \right) \tag{9.1}$$

corresponding to (6.9). This is derived from (7.19) in the same manner as (6.9) was from (4.5). As before, we have replaced the numbers $n_i - 1$ in (7.19) by n_i, an unessential change as far as the asymptotic theory is concerned. We recall that

$$r_i = n_i - m_{ii} \tag{9.2}$$

hence we need only show that the variables on the right are asymptotically normally distributed in order to have the same result for the r_i. Corresponding to Theorem 2 of Section 6, we state

THEOREM 1. *The variables*

$$\begin{aligned} x_{ij} - \frac{m_{ij} - np_i p_j}{\sqrt{n}} \qquad & i, j = 1, \cdots, k-1, \\ x_i = \frac{n_i - np_i}{\sqrt{n}} \qquad & i = 1, \cdots, k-1 \end{aligned} \tag{9.3}$$

are asymptotically normally distributed with zero means and variances and co-variances

$$\begin{aligned} \sigma_{ij,st} &= -3p_i p_j p_s p_t, & \sigma_{ij,it} &= -3p_i^2 p_j p_t, \\ \sigma_{ii,st} &= -3p_i^2 p_s p_t, & \sigma_{ii,ij} &= p_i^2 p_j (1 - 3p_i), \\ \sigma_{ij,ij} &= p_i p_j (1 - 3p_i p_j), & \sigma_{ii,ii} &= p_i^2 (1 + 2p_i - 3p_i^2), \\ \sigma_{ii,jj} &= -3p_i^2 p_j^2, & \sigma_{ij,s} &= -2p_i p_j p_s, \\ \sigma_{ii,s} &= -2p_i^2 p_s, & \sigma_{ij,i} &= p_i p_j (1 - 2p_i), \\ \sigma_{ii,i} &= 2p_i^2 (1 - p_i), & \sigma_{i,j} &= -p_i p_j, \\ & & \sigma_{i,i} = p_i(1 - p_i). & \end{aligned} \tag{9.4}$$

In these relations the symbols are defined by

$$\sigma_{ij,st} = \sigma_{x_{ij} x_{st}}, \qquad \sigma_{ij,s} = \sigma_{x_{ij} x_s}, \qquad \sigma_{i,j} = \sigma_{x_i x_j}$$

and different literal subscripts represent different numerical subscripts. These moments have been computed by means of the identity (6.12). The proof of the theorem is like that of Theorem 2 of Section 6 and will be omitted. We can now give the limiting form of the distribution of the r_i in (7.18) as

COROLLARY 1. *The variables*

$$x_i = \frac{r_i - np_i(1 - p_i)}{\sqrt{n}} \qquad i = 1, 2, \cdots, k \tag{9.5}$$

are asymptotically normally distributed with zero means and variances and covariances

$$\begin{aligned} \sigma_{ii} &= p_i(1 - p_i) - 3p_i^2(1 - p_i)^2, \\ \sigma_{ij} &= -p_ip_j(1 - 2p_i - 2p_j + 3p_ip_j). \end{aligned} \tag{9.6}$$

These limiting moments follow at once from equations (7.20).

COROLLARY 2. *The variable*

$$Q = \sum_1^k \sigma^{ij} x_i x_j$$

where the x_i are defined by (9.5) *and* $\| \sigma^{ij} \|$ *is the inverse of* (9.6), *is asymptotically distributed according to the* χ^2*-law with k degrees of freedom.*

COROLLARY 3. *If $r = \Sigma r_i$ denotes the total number of runs, then*

$$x = \frac{r - n(1 - \Sigma p_i^2)}{\sqrt{n}}$$

is asymptotically normally distributed with zero mean and variance

$$\sigma^2 = \Sigma p_i^2 + 2\Sigma p_i^3 - 3(\Sigma p_i^2)^2.$$

The author would like to record here his gratitude to Professor S. S. Wilks who suggested the problem and under whose direction this paper was written.

REFERENCES

[1] KARL PEARSON, *The Chances of Death and Other Studies in Evolution*, London, 1897, Vol. I, Chap. 2.
[2] KARL MARBE, *Naturphilosophische Untersuchungen zur Wahrscheinlichkeitslehre*, Leipzig, 1899.
[3] KARL MARBE, *Die Gleichförmigkeit in der Welt*, München, 1916.
[4] KARL MARBE, *Mathematische Bemerkungen*, München, 1916.
[5] KARL MARBE, *Grundfragen der angewandten Wahrscheinlichkeitsrechnung*, München, 1934.
[6] H. GRÜNBAUM, *Isolierte und reine Gruppen und die Marbe'sche Zahl "p"*, Würzburg, 1904.
[7] H. BRUNS, *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*, Leipzig, 1906, p. 216.
[8] L. v. BORTKIEWICZ, *Die Iterationen*, Berlin, 1917.
[9] R. v. MISES, *Zeit. f. angew. Math. u. Mech.*, Vol. 1 (1921), p. 298.
[10] E. ISING, *Zeit. f. Physik.*, Vol. 31 (1925), p. 253.
[11] J. WISHART AND H. O. HIRSHFELD, *London Math. Soc. Jour.*, Vol. 11 (1936), p. 227.
[12] W. L. STEVENS, *Annals of Eugenics*, Vol. 9 (1939), p. 10.
[13] A. WALD AND J. WOLFOWITZ, *Annals of Math. Stat.*, Vol. 11 (1940).
[14] J. A. JOSEPH, *Annals of Math. Stat.*, Vol. 10 (1939), p. 293.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

A GENERALIZATION OF THE LAW OF LARGE NUMBERS

By Hilda Geiringer

It is well known that the law of large numbers can be established for dependent as well as for independent chance variables by using Tchebycheff's inequality [1] and assuming that the variance of the sum of the variables tends towards infinity less rapidly than n^2.

In recent years v. Mises has introduced the notion of *statistical functions* [2] and has shown that, under certain assumptions the law of large numbers is still valid if, instead of the arithmetic mean of the n observations $x_1, \cdots, x_n$ a statistical function of these observations is considered. For example in the very special case, where the n collectives which have been observed are *identical* k-valued *arithmetic* distributions with probabilities $p_1, \cdots, p_k$ corresponding to the attributes $c_1, \cdots, c_k$ and with observed relative frequencies $n_1/n, \cdots, n_k/n$ one obtains the result: It is to be expected for every $\epsilon > 0$ with a probability P_n converging towards one as $n \to \infty$, that $|f(n_1/n, \cdots, n_k/n) - f(p_1, \cdots, p_k)| < \epsilon$ under very general conditions concerning the function f.

In the present paper we shall generalize these new results so that they will apply also to collectives which are not independent.

1. Lemma concerning alternatives. Let us consider the *n-dimensional collective* consisting of a *sequence of n trials* and let us assume that the n trials are alternatives, i.e. for each trial there are only two possible results which we denote by "success," "failure," by "occurrence," "non-occurrence" or by "1," "0." The total result of the n trials is expressed by n numbers each equal to 0 or 1. Let $v(x_1, x_2, \cdots, x_n)$ be the probability of obtaining the result x_1 at the first trial, x_2 at the second one, $\cdots$, x_n at the last one ($x_\nu = 0, 1$; $\nu = 1, \cdots, n$). In the same way we introduce $v_{12}(x, y) = \sum_{x_3, \cdots, x_n} v(x, y, x_3, \cdots, x_n)$ and generally $v_{\mu\nu}(x, y)$ as the probability that the μth result equals x, the νth equals y, ($\mu \neq \nu$), and finally let $v_\mu(x) = \sum_\nu v_{\mu\nu}(x, y)$ be the probability that the μth result equals x. In particular let us write

$$v_\mu(1) = p_\mu, \quad v_{\mu\nu}(1, 1) = p_{\mu\nu}, \quad (\mu, \nu = 1, \cdots, n; \mu \neq \nu)$$

p_μ being the probability of success in the μth trial and $p_{\mu\nu}$ the probability of simultaneous success both in the μth and νth trials.

The *variance* s_n^2 of the sum $(x_1 + \cdots + x_n)$ is easily found:

$$s_n^2 = \text{Var}(x_1 + \cdots + x_n) = \sum_{x_1,\cdots,x_n} (x_1 + \cdots + x_n - p_1 - \cdots - p_n)^2 v(x_1, \cdots, x_n)$$
$$= \sum_{x_1,\cdots,x_n} (x_1 - p_1)^2 v(x_1, \cdots, x_n) + \cdots$$
$$+ 2 \sum_{x_1,\cdots,x_n} (x_1 - p_1)(x_2 - p_2) v(x_1, \cdots, x_n) + \cdots$$
$$= \sum_{x_1} (x_1 - p_1)^2 v_1(x_1) + \cdots + 2 \sum_{x_1,x_2} (x_1 - p_1)(x_2 - p_2) v_{12}(x_1, x_2) + \cdots$$
$$= p_1(1 - p_1) + \cdots + p_n(1 - p_n) + 2(p_{12} - p_1 p_2) + \cdots + 2(p_{n-1,n} - p_{n-1} p_n).$$

Thus:

$$(1) \qquad s_n^2 = \text{Var}(x_1 + \cdots + x_n) = \sum_{\nu=1}^{n} p_\nu(1 - p_\nu) + 2 \sum_{\mu,\nu=1}^{n} (p_{\mu\nu} - p_\mu p_\nu).$$

The first sum on the right is $\leqq n/4$; the second one consists of $N = \frac{1}{2}n(n-1)$ terms, therefore we cannot be sure that it tends toward zero after division by n^2.

Putting $p_{\mu\nu} - p_\mu p_\nu = \alpha_{\mu\nu}^{(n)}$ we see immediately:

(*a*) *A necessary and sufficient condition for* $\lim_{n\to\infty} s_n/n = 0$ *is*

$$(2) \qquad \lim_{n\to\infty} 1/n^2 \sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)} = 0.$$

Denoting by σ_μ^2 the variance of $v_\mu(x)$ and by $r_{\mu\nu}$ the correlation coefficient of $v_{\mu\nu}(x, y)$ we have

$$\alpha_{\mu\nu}^{(n)} = p_{\mu\nu} - p_\mu p_\nu = r_{\mu\nu}\sigma_\mu\sigma_\nu .$$

We see that $\alpha_{\mu\nu}^{(n)}$ takes values between $-1/4$ and $+1/4$ and our conditions (2) postulates that the sum of these positive and negative terms tends towards infinity less rapidly than n^2. As to the meaning of the signs of these terms we see that a term $\alpha_{\mu\nu}^{(n)}$ will be $\gtreqless 0$, according as $p_{\mu\nu}/p_\nu \gtreqless p_\mu$. This means: the fact that the νth event has presented itself makes the occurrence of the μth event either more probable; or it is without influence on it; or it makes it less probable. And we see that s_n/n tends toward zero, only if there is a certain "equalization" or "stabilization" of positive and negative mutual influence. If in particular for a pair of values μ, ν, $r_{\mu\nu} = +1$, that is $v_{\mu\nu}(0, 1) = v_{\mu\nu}(1, 0) = 0$, the events must either both occur or both fail and $p_\mu = p_\nu$. If $r_{\mu\nu} = -1$ we have $v_{\mu\nu}(0, 0) = v_{\mu\nu}(1, 1) = 0$ the simultaneous occurrence is impossible and likewise the simultaneous failure, and $p_\mu + p_\nu = 1$. If we have $p_{\mu\nu} = 0$ (case of mutually exclusive events) then $p_\mu + p_\nu \leqq 1$.

Since $s_n^2 \geqq 0$ and $\sum_{\nu=1}^{n} p_\nu(1 - p_\nu) = \sum_{\nu=1}^{n} \sigma_\nu^2 \leqq n/4$ we conclude from (1) that $\sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)} \geqq -n/8$ and we obtain the following simple *sufficient* condition for the validity of (2):

(b) *Let us denote by m_n the number of all combinations $\mu, \nu(\mu \leqq n; \nu \leqq n; \mu \neq \nu)$, such that, however large n may be, $\alpha_{\mu\nu}^{(n)} > \epsilon$, where ϵ is a given positive number; then $\frac{1}{n^2} \sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)}$ converges toward zero if $\lim_{n\to\infty} m_n/n^2 = 0$.*

We have in fact

$$-\frac{n}{8} \leqq \sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)} \leqq m_n + (N - m_n)\epsilon$$

and dividing by n^2 we find that $\frac{1}{n^2} \sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)}$ is enclosed between $\frac{-1}{8n}$ and $m_n/n^2 + \epsilon \frac{N - m_n}{n^2}$ which both tend toward zero. Roughly speaking this condition implies that for "almost all" combinations of indices μ, ν, the $\alpha_{\mu\nu}^{(n)}$ converge toward "negative or vanishing correlation."

On the other hand the sum of all positive and negative terms in $\sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)}$ cannot become less than $-n/8$. Therefore, if "almost all" positive terms are supposed to tend towards zero it follows that also almost all negative terms tend toward zero. Thus we obtain the *sufficient* condition (c) which is neither more nor less general than (b):

(c) *The sum $\frac{1}{n^2} \sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)}$ tends towards zero as $n \to \infty$, if "almost all" the individual terms $\alpha_{\mu\nu}^{(n)} = p_{\mu\nu} - p_\mu p_\nu$ tend toward zero.* Or more exactly, the sum in question tends toward zero if $| \alpha_{\mu\nu}^{(n)} | \leqq \epsilon$ for every ϵ and sufficiently large n with the exception of μ_n terms where $\lim_{n\to\infty} \mu_n/n^2 = 0$. That is "convergence towards independence" for almost all combinations μ, ν of indices. Let us, for example, assume that all the p_ν are $\neq 0$ and all the $p_{\mu\nu} = 0$, then all the $\alpha_{\mu\nu}^{(n)}$ are certainly < 0 and (b) is fulfilled; but it is easily seen (3) that in this case $p_1 + p_2 + \cdots p_n \leqq 1$. Therefore all the products $p_\mu p_\nu$ (with the possible exception of a finite number) tend toward zero, and (c) holds as well.

2. Statistical functions. Suppose n observations have given the results $x_1, x_2, \cdots, x_n$. Let us assume for the sake of simplicity that they are all bounded between two real numbers A and B. To each real x corresponds the number $n\, S_n(x)$ of observations with a result $\leqq x$. $S_n(x)$ is a monotone non-decreasing step function with n steps, each of height $1/n$; however several steps may coincide at the same point. We have

(1) $$S_n(x) = 0 \quad \text{if} \quad x < A \quad \text{and} \quad S_n(x) = 1 \quad \text{if} \quad x \geqq B.$$

$S_n(x)$ is called by v. Mises the *partition* (Aufteilung) of the n observations. $S_n(x)$ coincides with the well known cumulative frequency distribution if the attributes $c_\kappa(\kappa = 1, \cdots k)$ and the corresponding relative frequencies $n_1/n, \cdots n_k/n$ are given.

A *statistical function* is a function of the $x_1, x_2, \cdots, x_n$ *which depends only on* $S_n(x)$, the partition of the n results. It will be denoted by $f\{S_n(x)\}$. If the c_κ and the n_κ/n are given then statistical function means simply "function of the relative frequencies" and it becomes a function of k variables. In $f\{S_n(x)\}$ the partition $S_n(x)$ takes the place of the independent variable. Such a statistical function has the following properties: (a) It is a symmetric function of the $x_1, x_2, \cdots, x_n$. That is, it is independent of the succession of the n results. (b) It is "homogeneous" in the following sense: If instead of n observations we have nl observations and if at the same time each x_ν is replaced by lx_ν then the statistical function is not changed.[1] Examples of statistical functions are the *moments*

$$\frac{1}{n}\sum_{\nu=1}^{n} x_\nu^r = \int x^r\, dS_n(x) = M_r^0$$

or, if $M_1^0 = \alpha$, the moments about the mean α:

$$\frac{1}{n}\sum_{\nu=1}^{n} (x_\nu - \alpha)^r = \int (x - \alpha)^r\, dS_n(x) = M_r, \text{ etc.}$$

The independent variable in $f\{S_n(x)\}$ is a partition; but in addition we shall define $f\{P(x)\}$ where $P(x)$ is a certain bounded distribution which is not necessarily a partition. A distribution $P(x)$ is called bounded if

$$(1') \qquad P(x) = 0 \quad \text{if} \quad x < A \quad \text{and} \quad P(x) = 1 \quad \text{if} \quad x \geqq B.$$

If this is true for a sequence $P_1(x)$, $P_2(x)$, $\cdots$ with the same A and B then the sequence is called *uniformly bounded.* Let us now consider a bounded partition $P(x)$ which in every point of continuity of $P(x)$ is the limit as $n \to \infty$ of a sequence of bounded partitions $S_n(x)$. *As* $S_n(x)$ *converges toward* $P(x)$, if $f\{S_n(x)\}$ converges towards a limit L which does not depend on the limiting process $S_n(x) \to P(x)$ then that limit shall be denoted by $f\{P(x)\}$; it will be called the *value of the statistical function at the "point"* $P(x)$ and $f\{S_n(x)\}$ will be called *continuous* at $P(x)$. The definition of continuity can be given also in the following way: Corresponding to every $\epsilon > 0$ exists an $\eta > 0$ such that

$$(2) \qquad | f\{S_n(x)\} - f\{P(x)\} | < \epsilon$$

for all values of n and for every bounded $S_n(x)$ *such that* at every point of continuity of $P(x)$

$$(3) \qquad | S_n(x) - P(x) | \leqq \eta.$$

In this case $f\{S_n(x)\}$ is called continuous at the point $P(x)$. Thus a statistical function is defined for bounded partitions and for certain bounded distributions which are not themselves partitions. If the continuity defined by (2) and (3) exists for a sequence $P_1(x)$, $P_2(x)$, $\cdots$ of bounded distributions with the same η

[1] This condition of homogeneity is fulfilled e.g. for $\sqrt{x_1x_2 \cdots x_n}$ but not for $x_1x_2 \cdots x_n$.

corresponding to a given ϵ, we call the statistical function *uniformly* continuous at the points $P_1(x)$, $P_2(x)$, $\cdots$.

3. The general law of large numbers. The generalization of the law of large numbers which we have in mind can be demonstrated in a way analogous to the demonstration given by v. Mises in the case of independent collectives if we introduce the results of paragraph 1 in order to estimate the variance. We shall consider here only one dimensional, bounded collectives in order to make clearer what is the essential of the generalization.

A sequence of dependent collectives $P_1(x)$, $P_2(x)$, $\cdots$, $P_n(x)$ can be given in the following manner. Let $P(x_1, x_2, \cdots, x_n)$ be the probability that the result of the first observation is $\leqq x_1$, of the second $\leqq x_2$, $\cdots$, of the nth $\leqq x_n$. This distribution will be said to be *bounded* in (A, B) if $P = 1$ when all the x_ν are $\geqq B$ and $P = 0$ if at least one of these arguments is less than A. From this n-dimensional distribution we deduce n one dimensional distributions

$$(1) \qquad \begin{aligned} &P_1(x) = P(x, B, \cdots, B), \\ &\qquad P_2(x) = P(B, x, B, \cdots, B), \cdots, P_n(x) = P(B, \cdots, B, x) \end{aligned}$$

where $P_\nu(x)$ is the probability that the νth observation be $\leqq x$. The $P_\nu(x)$ are uniformly bounded in (A, B) which is a consequence of $P(x_1, x_2, \cdots, x_n)$ having been assumed to be bounded in this interval. In an analogous way we deduce from $P(x_1, x_2, \cdots, x_n)$ the $\frac{1}{2}n(n-1)$ uniformly bounded two dimensional distributions

$$(2) \quad P_{12}(x, y) = P(x, y, B, \cdots B), \qquad P_{13}(x, y) = P(x, B, y, B, \cdots B), \cdots .$$

Here $P_{\mu\nu}(x, y)$ is the probability that the μth result is $\leqq x$, the νth result $\leqq y$, and we have $P_{\mu\nu}(x, y) = P_{\nu\mu}(y, x)$. Of course we have also

$$(1') \qquad \begin{aligned} P_1(x) &= P_{12}(x, B) = P_{13}(x, B) = \cdots = P_{1n}(x, B) \\ P_2(x) &= P_{12}(B, x) = P_{23}(x, B) = \cdots = P_{2n}(x, B) \text{ etc.} \end{aligned}$$

If we put in (2) $x = y$ we obtain $P_{\mu\nu}(x, x) = P_{\nu\mu}(x, x)$ and we introduce

$$(3) \qquad P_{\mu\nu}(x, x) = P_{\nu\mu}(x) = P_{\mu\nu}(x)$$

the probability that both the μth and the νth observation is $\leqq x$. Then $P_{\mu\nu}(x)$ equals zero if $x < A$ and equals one if $x \geqq B$, and this is valid with the same A and B for all the distributions $P_{\nu\mu}(x)$.

Now if $p_1, p_2, \cdots, p_n$ are the probabilities of success for n general alternatives Tchebycheff's Lemma asserts that the probability W that the average $(x_1 + x_2 + \cdots + x_n)/n$ of n observations differs by more than η from its expectation $(p_1 + p_2 + \cdots + p_n)/n$ is subject to the following inequality

$$(4) \qquad W \leqq \frac{1}{\eta^2} \operatorname{Var}\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) = \frac{s_n^2}{\eta^2 n^2}.$$

Here s_n^2 is given by (1) of paragraph 1.

Let us introduce the average $\bar{P}_n(x)$ of the $P_\nu(x)$:

$$\bar{P}_n(x) = [P_1(x) + P_2(x) + \cdots + P_n(x)]/n \tag{5}$$

and let Q_n be the probability that at any point of continuity of $\bar{P}_n(x)$ the inequality

$$| S_n(x) - \bar{P}_n(x) | > \eta \tag{6}$$

holds. Our aim will be to show that for every η under certain restrictions regarding the given collectives, Q_n tends toward zero as n tends toward infinity.

For a fixed point x' the probabilities $P_\nu(x) = p_\nu$ and $P_{\mu\nu}(x) = p_{\mu\nu}$ are constants and we put $\bar{P}_n(x) = \bar{p}_n = (p_1 + p_2 + \cdots p_n)/n$. The probability that in x'

$$| S_n(x') - \bar{P}_n(x') | > \eta/2 \tag{7}$$

is then, according to (4) smaller than $(s_n^2)_{x'}/(\frac{1}{2}\eta)^2 n^2$. Here we denote by $(s_n^2)_{x'}$ the value of s_n^2 in x' (as given by (1) in paragraph 1).

Now we divide the interval (A, B) in N parts in such a way that in every one of the N intervals e.g. in (x', x'') the variation

$$\delta = \bar{P}_n(x'') - \bar{P}_n(x') \leqq \eta/2. \tag{8}$$

If there is at x' (or at x'') a step of $\bar{P}_n(x)$ we take the limit which $\bar{P}_n(x)$ approaches as $x \to x'$ (or x'') from the interior of the interval. In order to obtain such a division we need only divide the total variation 1 of $\bar{P}_n(x)$ in $2/\eta$ equal parts and project these points of division on $\bar{P}_n(x)$, disposing however in a suitable way of horizontal parts of $\bar{P}_n(x)$. The abscissae of these points form the endpoints of the N intervals. If there is a step of $\bar{P}_n(x)$ at an endpoint of one of these intervals the variation in both the adjacent intervals can only be diminished. It is further possible that the two ends of an interval coincide $x' = x''$, this will be so if $\bar{P}_n(x)$ has for x' a step $> \eta/2$. In any case we have a division in $N \leqq 2/\eta$ intervals such that all the points of continuity of $\bar{P}_n(x)$ are enclosed in them and in each of these intervals (8) is valid.

Let us now assume that in the left end point x' of the rth interval (x', x'') the inequality

$$| S_n(x') - \bar{P}_n(x') | \leqq \eta/2 \tag{9}$$

is valid. Then we have for *every* x between x' and x''

$$| S_n(x) - \bar{P}_n(x) | \leqq \eta/2 + \delta \leqq \eta. \tag{10}$$

Because, since $S_n(x)$ and $\bar{P}_n(x)$ are both monotone, the difference $S_n(x') - \bar{P}_n(x')$ cannot increase by more than $\delta \leqq \eta/2$ as x varies from x' to x''. Therefore if (6) is valid for any point x in this interval then (7) must be valid for the left end point x' of this interval and the probability q_r of this latter inequality is less than or equal to $4(s_n^2)_{x'}/\eta^2 n$.

But there are N intervals with the left endpoints x_1', x_2', $\cdots x_N'$ and the

probability that (6) may be valid in any point belonging to any one of these intervals is $\leqq q_1 + q_2 + \cdots + q_N$. Denoting by s_n^2 the greatest of the N variances $(s_n^2)_{x_1'}, (s_n^2)_{x_2'}, \cdots, (s_n^2)_{x_N'}$ we have for Q_n (which is the probability that (6) may be valid at any point of continuity of $P(x)$) the inequality

$$Q_n \leqq q_1 + q_2 + \cdots + q_N \leqq \frac{4N}{\eta^2 n^2} s_n^2 \leqq \frac{8}{\eta^3} \frac{s_n^2}{n^2}. \tag{11}$$

Therefore Q_n tends toward zero for every η if s_n/n tends toward zero.

But according to (2) in paragraph 1, s_n/n tends toward zero if for every x in (A, B)

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{\mu,\nu=1}^{n} [P_{\mu\nu}(x) - P_\mu(x)P_\nu(x)] = 0. \tag{12}$$

Considering the definition of continuity of a statistical function we have obtained the following result:

As in (1′), (2), (3) *and* (5) *let* $P_{\mu\nu}(x, y)$ *be two dimensional distributions* ($\mu, \nu = 1, \cdots, n; \mu \neq \nu$), *uniformly bounded in* (A, B); $P_{\mu\nu}(x, B) = P_\mu(x)$; $P_{\mu\nu}(x, x) = P_{\mu\nu}(x)$ *and* $\bar{P}_\nu(x) = 1/\nu(P_1(x) + P_2(x) + \cdots + P_\nu(x))$.

If the variable partition $S_n(x)$ *is bounded in* (A, B) *and if* $f\{S_n(x)\}$ *is uniformly continuous at the "points"* $\bar{P}_1(x), \bar{P}_2(x), \cdots$ *then the probability that*

$$|f\{S_n(x)\} - f\{\bar{P}_n(x)\}| > \epsilon \tag{13}$$

tends toward zero for every ϵ *as* $n \to \infty$, *provided* (12) *is uniformly valid for every* x *in* (A, B).

4. Examples. Let us illustrate by simple examples.

1) In order to define the $P_\nu(x)$ etc. mentioned in our theorem we define the n-dimensional distribution $P(x_1, x_2, \cdots x_n)$ used at the beginning of paragraph 3 by indicating the probability density

$$\begin{aligned} \mu(x_1, x_2, \cdots, x_n) &= C_n[1 - x_1x_2 \cdots x_n] \quad \text{in the "unit cube"}, \\ &= 0 \quad \text{elsewhere}. \end{aligned} \tag{1}$$

The corresponding probability distribution is

$$P(x_1, x_2, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \mu(x_1, x_2, \cdots, x_n)\, dx_1 \cdots dx_n. \tag{2}$$

By putting

$$C_n = \frac{2^n}{2^n - 1}, \tag{3}$$

we see that $P(x_1, x_2, \cdots, x_n)$ equals unity if all the arguments are $\geqq 1$ and it equals zero if one of these arguments is less than 0. Therefore $P(x_1, x_2, \cdots, x_n)$ is bounded in the unit cube.

From (1) we deduce the two-dimensional densities

$$(4)\qquad \begin{aligned} v_{\mu\nu}(x, y) &= C_n\left(1 - \frac{xy}{2^{n-2}}\right) && \text{in the unit square,} \\ &= 0 && \text{elsewhere} \end{aligned}$$

and the distributions

$$(5)\qquad P_{\mu\nu}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} v_{\mu\nu}(x, y)\, dx\, dy.$$

We see that

$$\begin{aligned} P_{\mu\nu}(x, y) &= C_n xy\left(1 - \frac{xy}{2^n}\right) && \text{in the unit square} \\ &= 0 && \text{if } x \text{ or } y \leqq 0 \\ &= 1 && \text{if } x \text{ and } y \geqq 1 \end{aligned}$$

and e.g. for $x \geqq 1$, $0 < y < 1$ we have $P_{\mu\nu}(x, y) = P_{\mu\nu}(1, y)$ etc. Thus the $P_{\mu\nu}(x, y)$ are completely given.

It follows from (3) that $-C_n/2^n = 1 - C_n$; therefore putting $C_n = C$ we have in (0, 1)

$$(6)\qquad \begin{aligned} P_{\mu\nu}(x, x) = P_{\mu\nu}(x) &= Cx^2 + (1 - C)x^4 \\ P_\nu(x) &= Cx + (1 - C)x^2 \end{aligned}$$

therefore

$$(7)\qquad P_{\mu\nu}(x) - P_\mu(x)P_\nu(x) = C(1 - C)x^2(1 - x)^2$$

is < 0 for every x in (0, 1) since $C > 1$. For $x \leqq 0$, $P_{\mu\nu}(x)$ and $P_\nu(x)$ both equal zero and for $x \geqq 1$ they both equal 1. Therefore our conditions of paragraph 1 are fulfilled. We see that C_n tends towards unity as $n \to \infty$, therefore for every x in (0, 1) $P_{\mu\nu}(x) - P_\mu(x)P_\nu(x)$ tends towards zero, we have "convergence towards independence" but by no means independence.

This example was based on a *symmetric density*. Let us give an example of asymmetric and *arithmetic* distributions. For the sake of simplicity let $P_1(x)$, $P_2(x)$, $\cdots$ be arithmetic distributions each with only three steps at $x = 0$, 1 and 2. As starting point we take the n-dimensional arithmetic distribution $v(x_1, x_2, \cdots x_n)$ which gives the probability that the first result equals x_1, the second x_2, $\cdots$, the nth x_n, the x_ν being equal to 0 or 1 or 2; thus $v(x_1, x_2, \cdots, x_n)$ takes 3^n values the sum of which equals unity. We deduce the two dimensional distributions $v_{\mu\nu}(x, y)$, e.g. $v_{12}(x, y) = \sum_{x_3, \cdots, x_n} v(x, y, x_3, \cdots, x_n)$, the probability that the first result equals x, the second y, and finally the $v_1(x) = \sum_y v_{12}(x, y)$, etc. According to the definitions of $P_\nu(x)$ and $P_{\mu\nu}(x)$ we have then:

$$(8)\qquad \begin{aligned} P_\nu(x) &= 0 && (x < 0)\\ &= v_\nu(0) && (0 \leqq x < 1)\\ &= v_\nu(0) + v_\nu(1) && (1 \leqq x < 2)\\ &= 1 && (2 \leqq x),\end{aligned}$$

$$(9)\qquad \begin{aligned} P_{\mu\nu}(x) &= 0 && (x < 0)\\ &= v_{\mu\nu}(0, 0) && (0 \leqq x < 1)\\ &= v_{\mu\nu}(00) + v_{\mu\nu}(10) + v_{\mu\nu}(01) + v_{\mu\nu}(11) && (1 \leqq x < 2)\\ &= 1 && (2 \leqq x).\end{aligned}$$

Now we subject $v(x_1, \cdots, x_n)$ to the following conditions: Every $v(x_1, \cdots, x_n)$ equals zero if it contains either: at least two "zeros," or: at least one "zero" and one "one," or: at least two "ones." All the other v-values are supposed to be different from zero. Then we have

$$v_{\mu\nu}(0, 0) = v_{\mu\nu}(1, 0) = v_{\mu\nu}(0, 1) = v_{\mu\nu}(1, 1) = 0$$

therefore $P_{\mu\nu}(x) = 0$ for $x < 2$ and $P_{\mu\nu}(x) = 1$ for $x \geqq 2$. On the other hand $v_\nu(0) = v(2, 2, \cdots 2, 0, 2, \cdots 2)$ and $v_\nu(1) = v(2, 2, \cdots 2, 1, 2, \cdots 2)$ therefore $P_\nu(x) \neq 0$ for $0 \leqq x < 2$ and we have thus for every finite n

$$\begin{aligned} P_{\mu\nu}(x) - P_\mu(x)P_\nu(x) &= 0 \quad \text{for } x < 0 \text{ and } x \geqq 2,\\ &< 0 \quad \text{for } 0 \leqq x < 2.\end{aligned}$$

Therefore the condition (b) of paragraph 1 is fulfilled and thus (12) paragraph 3 holds.

I hope to have the opportunity to discuss more general applications of this theorem later.

A generalization of the *strong* law of large numbers may be given in a similar way.

REFERENCES

[1] B. H. Camp, *The Mathematical Part of Elementary Statistics*, New York, 1934, page 256.
[2] R. v. Mises, "Die Gesetze der großen Zahl für statistische Funktione," *Monatshefte für Math. u. Physik*, 1936, p. 105–128.
[3] H. Geiringer, "Sur les variables aléatoires arbitrairement liées," *Revue de l'Union Interbalcanique*, 1938, p. 6.

Bryn Mawr College,
Bryn Mawr, Pennsylvania.

CONDITIONS FOR UNIQUENESS IN THE PROBLEM OF MOMENTS

By M. G. Kendall

It was shown by Stieltjes [1] that in some circumstances it is possible for two different frequency distributions to have the same set of moments. For instance, the integral

$$\int z^{4n+3} e^{-z} e^{iz}\, dz$$

around a contour consisting of the positive x-axis, the infinite quadrant and the positive y-axis is seen to be zero and it follows that

$$\int_0^\infty x^n e^{-x^{\frac{1}{4}}} \sin x^{\frac{1}{4}}\, dx = 0.$$

Thus the frequency distribution

$$dF = \tfrac{1}{6} e^{-x^{\frac{1}{4}}}(1 - \lambda \sin x^{\frac{1}{4}})\, dx \qquad \begin{array}{l} 0 \leq x \leq \infty, \\ 0 \leq \lambda \leq 1 \end{array} \tag{1}$$

has moments which are independent of λ, and equation (1) may be regarded as defining a whole family of distributions each of which has the same moments. It is easy to see that moments of all orders exist, and in fact

$$\mu_r' \text{ (about the origin)} = \tfrac{1}{6}(4r + 3)!.$$

A second example of the same kind, also due to Stieltjes, is the distribution

$$dF = \frac{1}{e^{\frac{1}{4}}\sqrt{\pi}} x^{-\log x}\{1 - \lambda \sin (2\pi \log x)\}\, dx \qquad \begin{array}{l} 0 \leq x \leq \infty, \\ 0 \leq \lambda \leq 1, \end{array} \tag{2}$$

for which

$$\mu_r' = e^{\frac{1}{4}r(r+2)}.$$

The question naturally arises, what are the conditions under which a given set of moments determines a frequency distribution uniquely? The question is of great interest to mathematicians, being closely linked with problems in the theory of asymptotic series, continued fractions and quasi-analytic functions; and it also has importance for statisticians since there is sometimes occasion to be satisfied that a problem of finding a frequency distribution has been uniquely solved by the ascertainment of its moments or semi-invariants. Stieltjes himself considered a more general problem: given a set of constants c_0,

$c_1, \cdots c_r, \cdots$ does there exist a function F, non-decreasing and possessing an infinite number of points of increase, such that

$$\int_0^\infty x^r \, dF = c_r \tag{3}$$

and under what conditions is F unique, except for an additive constant? Stieltjes showed that if we express the series

$$\sum_{r=0}^{\infty} (-1)^r \frac{c_r}{z^r} \tag{4}$$

as a continued fraction of the form

$$\frac{1}{a_1 z +} \; \frac{1}{a_2 +} \; \frac{1}{a_3 z +} \; \frac{1}{a_4 +} \cdots \; \frac{1}{a_{2n-1} z +} \; \frac{1}{a_{2n} +} \cdots \tag{5}$$

it is a necessary and sufficient condition for the existence of at least one F that all the a's be positive; and that the function is unique or not according as the series $\sum_{r=0}^{\infty} (a_r)$ diverges or converges. (If the a's are positive it must do one or the other.) The integral of equation (3) is to be interpreted in the general Stieltjes sense, so that the result applies to discontinuous as well as to continuous distributions. This is also true of the results obtained below.

Hamburger [2] discussed the similar problem when the limits of the integral in equation (3) are $\pm \infty$, and showed that a function F exists if the expression of (4) as a continued fraction of the form

$$\frac{b_0}{a_0 + z +} \; \frac{b_1}{a_1 + z +} \; \frac{b_2}{a_2 + z +} \cdots$$

gives positive values of the b's. In order that F may be unique it is necessary and sufficient that the continued fraction be completely (vollständig) convergent in the sense defined by Hamburger.

Unfortunately these criteria, though mathematically complete, are not very useful to statisticians because as a rule it is too difficult to express the coefficients a and b explicitly enough in terms of the given c's to enable questions of sign or of convergence to be decided. So far as I know, no more convenient criterion for the general Stieltjes problem has been found; but progress is possible if one considers the narrower question: given a set of moments, is the distribution which furnished them unique, that is to say, can any other distribution have furnished them? This is more limited than the Stieltjes problem because we know that at least one solution exists.

Contributions to this subject have been made by Lévy [3] and Carleman [4]. Lévy shows that if moments of all orders exist and are positive it is a sufficient condition for them to determine a distribution uniquely that $\mu_n^{1/n}/n$ remains finite as n tends to infinity. (Here and elsewhere in this paper μ_r refers to the moment of order r about any point, not necessarily the mean.) Carleman shows

that, for the case of limits $-\infty$ to $+\infty$ the moments determine the distribution uniquely if

$$\sum_{r=0}^{\infty} \frac{1}{(\mu_{2r})^{1/(2r)}}$$

diverges. For the limits 0 to ∞ he gives the corresponding series

$$\sum_{r=0}^{\infty} \frac{1}{(\mu_r)^{1/(2r)}}$$

a criterion which can be improved upon, as will be shown below.

The purpose of this paper is to develop criteria of this kind more systematically and to give more general criteria suitable in cases where the moments are not known explicitly but the behavior of the frequency distribution at its terminals is known.

Three preliminary points necessary for the later argument may be noted.

(1) Define the absolute moment of order r by

$$\nu_r = \int_{-\infty}^{\infty} |x^r| \, dF$$

and recall that

$$\nu_1 \leq \nu_2^{\frac{1}{2}} \leq \nu_3^{\frac{1}{3}} \leq \cdots \leq \nu_r^{1/r} \leq \cdots$$

(cf. Hardy and others, [5]). In other words the quantities $\nu_r^{1/r}$ form an increasing positive sequence and their reciprocals a decreasing positive sequence.

(2) The quantity $\nu_n^{1/n}/n$ must either tend to a limit or diverge to infinity as $n \to \infty$. For suppose that

$$\overline{\lim}\, \nu_n^{1/n}/n = k,$$

$$\underline{\lim}\, \nu_n^{1/n}/n = l.$$

Writing temporarily $\nu_n^{1/n} = a_n$, we have that, given ϵ there is an N such that

$$a_n/n > k - \epsilon$$

for an infinity of values of n greater than N. Similarly there is an M such that

$$a_n/n < l + \epsilon$$

for an infinity of values of n greater than M. Now choose ρ such that a_ρ, $a_{\rho+1}$ are two consecutive values, one near the upper limit and one near the lower limit. This can always be done and we can take ρ as large as we please. We then have

$$a_\rho > \rho(k - \epsilon)$$

$$a_{\rho+1} < (\rho + 1)(l + \epsilon)$$

and hence, since $a_{\rho+1} \geq a_\rho$

$$(k - \epsilon)\rho < (\rho + 1)(l + \epsilon)$$

giving

$$(k - l) < \frac{l}{\rho} + 2\epsilon + \frac{\epsilon}{\rho}.$$

Thus $k - l$ can be made as small as we please and is thus zero.

The argument can be very simply adapted to the case in which k is infinite, and if l is not finite k, being not less than l, is infinite. Thus as $n \to \infty$ either $\lim a_n/n$ exists or $a_n/n \to \infty$.[1]

(3) If any moment fails to converge, so will all moments of higher order. It is evident that more than one distribution can exist having a limited number of finite moments given and the remainder infinite. Thus we need only consider the case when moments of all orders exist. Furthermore, if any even moment exists the absolute moment of next lowest order must exist; for if $\int_{-\infty}^{\infty} x^{2n}\,dF$ exists, then each of $\int_{-\infty}^{0} x^{2n}\,dF$ and $\int_{0}^{\infty} x^{2n}\,dF$ exist separately, each being positive. Hence $\int_{-\infty}^{0} x^{2n-1}\,dF$ and $\int_{0}^{\infty} x^{2n-1}\,dF$ exist separately and thus $\int_{-\infty}^{\infty} |x^{2n-1}|\,dF = -\int_{-\infty}^{0} x^{2n-1}\,dF + \int_{0}^{\infty} x^{2n-1}\,dF$ exists. Hence we need only consider the case in which absolute moments of all orders exist.

THEOREM 1. *A set of moments determines a distribution uniquely if the series* $\sum_{r=0}^{\infty} \frac{\nu_r t^r}{r!}$ *converges for some real non-zero* t.

Consider the characteristic function

$$\phi(t) = \int_{-\infty}^{\infty} e^{ixt}\,dF.$$

This is uniformly continuous in t, and so are its derivatives of all orders. Thus we have, in the neighborhood of $t = 0$ the Maclaurin expansion

$$\phi(t) = \sum_{r=0}^{r} \left\{ \frac{t^r}{r!} \left[\frac{d^r \phi}{dt^r} \right]_{t=0} \right\} + R$$

$$= \sum_{r=0}^{r} \frac{(it)^r}{r!} \mu_r + R.$$

[1] This proof is necessary to the use of limits in the following theorems, but Theorems 2 and 3 are equally valid if $\overline{\lim}$ is substituted for lim therein. It is not generally true that if a_n and b_n are increasing monotonic sequences either $\lim a_n/b_n$ exists or $a_n/b_n \to \infty$ as $n \to \infty$.

Consequently, under the condition of the theorem, which implies that $\Sigma \frac{(it)^r}{r!} \mu_r$ is absolutely convergent for some radius ρ, $\phi(t)$ has a Taylor expansion in the neighborhood of the origin and is thus uniquely determined by the moments for $t < \rho$. Furthermore, in the neighborhood of $t = t_0$ we have

$$\phi(t) = \sum_{r=0}^{r} \left\{ \frac{i^r (t - t_0)^r}{r!} \int_{-\infty}^{\infty} x^r e^{it_0 x} \, dF \right\} + R.$$

The modulus of the coefficient of $\frac{(t - t_0)^r}{r!}$ is not greater than ν_r. Therefore $\phi(t)$ can be expanded in the neighborhood of $t = t_0$ in a Taylor series with a radius of convergence at least equal to ρ. Hence the function defining $\phi(t)$ in the neighborhood of the origin can be continued analytically throughout the range $-\infty$ to $+\infty$ and $\phi(t)$ is uniquely determined in that range.

But the characteristic function unqiuely determines the distribution; and hence the theorem follows.

As a result of Theorem 1 we have the following generalization of the criterion given by Lévy.

THEOREM 2. *A set of moments completely determines a distribution if* $\lim_{n \to \infty} \nu_n^{1/n}/n$ *is finite.*

It has already been seen that unless $\nu_n^{1/n}/n$ becomes infinite the limit exists. By the Cauchy test for convergence the series $\Sigma \frac{\nu_r t^r}{r!}$ converges if

$$\lim_{n \to \infty} \left(\frac{\nu_n t^n}{n} \right)^{1/n} < 1. \tag{7}$$

As $n \to \infty$, $(n!)^{1/n}$ tends, in accordance with Stirling's theorem, to $(\sqrt{2\pi n}\, e^{-n} n^n)^{1/n}$ i.e. to n/e. Consequently the condition (7) becomes

$$\lim\, [\nu_n^{1/n}/n]\, et < 1.$$

Thus if $\lim \nu_n^{1/n}/n = k$, say, the inequality (7) is satisfied for $t < 1/(ek)$ and the theorem follows.

An important corollary, which enables us to disregard the absolute moments (which may not be given if part of the range is negative) is

THEOREM 3. *A set of moments uniquely determines a distribution if* $\lim_{n \to \infty} \mu_{2n}^{1/(2n)}/n$ *is finite.*

For
$$\nu_{2n-1}^{1/(2n-1)} \leq \nu_{2n}^{1/(2n)} = \mu_{2n}^{1/(2n)}.$$

Thus,
$$\lim \frac{1}{2n-1} \cdot \nu_{2n-1}^{1/(2n-1)} \leq \lim \frac{2n}{2n-1} \cdot \frac{1}{2n} \mu_{2n}^{1/(2n)}$$
$$\leq \lim \frac{1}{2n} \mu_{2n}^{1/(2n)}$$

and is therefore finite if the limit on the right is finite. Thus lim $\nu_n^{1/n}/n$, which cannot be greater than the greater of the two limits of $\nu_{2n-1}^{1/(2n-1)}/(2n-1)$ and $\nu_{2n}^{1/(2n)}/(2n)$, must be finite; and the theorem follows from Theorem 2.

Now consider the series $\sum_{r=0}^{\infty} \frac{1}{\nu_r^{1/r}}$. Since the successive terms form a monotonic sequence it is a sufficient as well as a necessary condition for convergence that $n/\nu_n^{1/n}$ tend to zero. Thus, if the series is divergent $n/\nu_n^{1/n}$ cannot tend to zero and so $\nu_n^{1/n}/n$ cannot become infinite. Hence it must tend to a finite limit, which may in particular be zero. Hence from Theorem 3 we get

THEOREM 4. *A frequency distribution is uniquely determined by its moments if* $\sum_{r=0}^{\infty} \frac{1}{\nu_r^{1/r}}$ *diverges.*

Since $1/\nu_r^{1/r}$ is a decreasing sequence the series $\Sigma\, 1/\nu_r^{1/r}$ converges or diverges with $\Sigma\, 1/\mu_{2r}^{1/(2r)}$. The Carleman criterion, given by him for the case of limits $\pm\infty$, follows. For the case of limits 0 to ∞ the absolute moments are the same as the moments and the criterion can be the divergence of either $\Sigma\, 1/\mu_r^{1/r}$ or $\Sigma\, 1/\mu_{2r}^{1/(2r)}$. Since μ_r is greater than unity in the type of case under consideration the former series provides a more stringent test than that given by Carleman.

At first sight it is rather surprising that the uniqueness of the distribution depends only on the behavior of the even moments, particularly when, by a simple extension of the above result, it is seen that a sufficient condition for uniqueness is the divergence of $\Sigma\, 1/\mu_{4n}^{1/(4n)}$ or $\Sigma\, 1/\mu_{mn}^{1/(mn)}$ or any infinite subset chosen from the moments. It will, however, be remembered that the odd moments are conditioned to some extent by the even moments, and that uniqueness is really determined by the limiting form of ν_n as n tends to infinity.

It is evident that other tests may be derived from Theorem 1 by using the various tests for the convergence of an infinite series. For instance it is a sufficient condition for a set of moments to determine uniquely a distribution with positive range that

$$\frac{\mu_n}{n!}\Big/\frac{\mu_{n+1}}{(n+1)!} = 1 + \frac{\alpha}{n} + O\left(\frac{1}{n^{1+\beta}}\right), \quad \text{where } \begin{matrix}\alpha > 1\\ \beta > 0\end{matrix}$$

i.e. that

$$\frac{\mu_n}{\mu_{n+1}} = 1 + \frac{\gamma}{n} + O\left(\frac{1}{n^{1+\beta}}\right), \qquad \gamma > 0. \tag{8}$$

It may be noted in passing that the distribution

$$dF = e^{-x}\,dx \qquad 0 \le x \le \infty,$$

for which

$$\mu_r \text{ (about origin)} = r!$$

is completely determined by its moments. In fact, by direct reference to Theorem 1 we see that the series $\Sigma\,(it)^r$ converges for $t < 1$.

A frequency distribution of finite range is uniquely determined by its moments. For if the range is 0 to A we have

$$\mu_r = \int_0^A x^r \, dF \leq A^r$$

and hence $1/\mu_r^{1/r} \geq 1/A$ so that the series $\Sigma\, 1/\mu_r^{1/r}$ is divergent.

A proof for the case when the frequency distribution is continuous has been given by Lévy, though on entirely different lines from the above.

THEOREM 5. *A frequency distribution of infinite range is uniquely determined by its moments if it tends to zero at the infinite terminals faster than* e^{-x}.

Consider first of all the case when only one end of the range is infinite, so that we may take the range to be 0 to ∞.

If $(\mu_n/n!)^{1/n}$ has a finite limit the distribution is unique, by Theorem 2. We have then only to consider the cases (if any) in which $(\mu_n/n!)^{1/n}$ tends to infinity. It will be shown that in fact such cases do not occur.

Given any (small) ϵ there exists an X such that

$$\frac{f(x)}{e^{-x}} < \epsilon, \qquad x > X$$

where $f(x)$ is the distribution. Thus

$$\int_X^\infty f(x) x^n \, dx < \epsilon \int_X^\infty e^{-x} x^n \, dx < \epsilon n!. \tag{9}$$

This is true for all n and X is independent of n. Now,

$$\int_0^\infty f(x) x^n \, dx = \int_0^X f(x) x^n \, dx + \int_X^\infty f(x) x^n \, dx.$$

The first integral on the right is not greater than X^n. The integral on the left tends, for large n, to something of greater order than $n!$, by our hypothesis, and hence to something of greater order than n^n. This is of greater order than X^n (since X, however large, is independent of n) and consequently the second integral on the right is also of greater order than $n!$. But this is contrary to equation (9).

The case for the range which is infinite in both directions may be dealt with similarly.

It is easily seen that the two examples of equations (1) and (2) do not tend to infinity faster than e^{-x}.

Except for the general result of Stieltjes, all the above criteria provide sufficient conditions, but whether the condition of Theorem 1 is also necessary is not certain. An inquiry into the circumstances in which the moment-series of Theorem 1 does not converge throws some light on the question.

It will be remembered that the characteristic function always exists and is uniformly continuous in t. Since the moments of all orders are assumed to exist we always have

$$\left[\frac{d^r}{dt^r}\phi(t)\right]_{t=0} = (i)^r\mu_r.$$

Thus, if $\phi(t)$ can be expanded in an infinite Taylor series that series must be $\Sigma \frac{(it)^r}{r!}\mu_r$. And if this series does not converge then $\phi(t)$ cannot be expanded as an infinite Taylor series. But it can always be expanded in the finite form with remainder

$$\phi(t) = \sum_{r=0}^{r} \frac{(it)^r}{r!}\mu_r + R.$$

Thus, when the series does not converge, $\phi(t)$ can be expanded in powers of t only asymptotically.

Now it is known that there exist an infinite number of functions which have a given set of coefficients in an asymptotic expansion; for instance, if $\psi(t)$ has an asymptotic expansion in t the functions $\psi(t) + \lambda t^{-\log t}$ all have the same expansion. It is therefore hardly surprising that when the conditions of Theorem 1 break down there can be more than one frequency distribution with the same set of moments.

But it does not follow from what has been said that there *must* be more than one frequency distribution. There must be more than one function, but those functions may not qualify as frequency distributions, e.g. they may be negative in part of the range. In the example just given $t^{-\log t}$ cannot be a characteristic function, for it does not obey the well-known condition that $\phi(t)$ and $\phi(-t)$ should be conjugate.

However, the question is more of mathematical than of statistical interest since the criteria provided above are likely to be adequate for the distributions encountered in practice. For example they establish the uniqueness of the Pearson curves (including the normal curve), the Poisson and the binomial. It would seem that distributions like those of equations (1) and (2) will appear only as statistical curiosities.

REFERENCES

[1] J. Stieltjes, *Recherches sur les fractions continues*, Oeuvres, Groningen, 1918.
[2] H. Hamburger, "Über eine Erweiterung des Stietjesschen Momentproblems," *Math. Annalen*, Vol. 81 (1920), p. 235, and Vol. 82 (1921), pp. 120 and 168.
[3] P. Lévy, *Calcul des Probabilités*, 1925, Paris.
[4] T. Carleman, *Les Fonctions Quasi-analytiques*, 1925, Paris.
[5] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, 1934, Cambridge, England.

London, England.

ON SAMPLES FROM A NORMAL BIVARIATE POPULATION

By C. T. Hsu

1. Introduction. In a number of papers written during the last ten years, J. Neyman and E. S. Pearson[1] have discussed certain general principles underlying the choice of tests of statistical hypotheses. They have suggested that any formal treatment of the subject requires in the first place the specification of (*i*) the hypothesis to be tested, say H_0, (*ii*) the admissible alternative hypotheses. An appropriate test will then consist of a rule to be applied to observational data, for rejecting H_0 in such a way that (*iii*) the risk of rejecting H_0 when it is true is fixed at some desired value (e.g., 0.05 or 0.01), (*iv*) the risk of failing to reject H_0 when some one of the admissible alternatives is true is kept as small as possible. With these general principles in mind, they have investigated how best the condition (*iv*) may be satisfied in different classes of problems. In many cases, though not in all, it has been found that the conditions are satisfied by the test obtained from the use of what has been termed the likelihood ratio, [9], [10], [14]. Once the problem has been specified, the test criterion is usually very easily found, although its sampling distribution, if H_0 is true, often presents great difficulties. In the present paper, I propose to use this method to obtain appropriate tests for a number of hypotheses concerning two normally correlated variables. The investigation was suggested by a recent application of the method by W. A. Morgan [6] to a problem originally discussed by D. J. Finney [3].

2. The hypotheses and the appropriate criteria. A sample of two variables x_1 and x_2 is supposed to have been drawn at random from a normal bivariate population, with the distribution

$$
\begin{aligned}
p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp\left\{-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\xi_1}{\sigma_1}\right)^2\right.\right. \\
\left.\left. - 2\rho_{12}\left(\frac{x_1-\xi_1}{\sigma_1}\right)\left(\frac{x_2-\xi_2}{\sigma_2}\right) + \left(\frac{x_2-\xi_2}{\sigma_2}\right)^2\right]\right\}
\end{aligned}
\tag{1}
$$

where ξ_1, ξ_2, σ_1, σ_2, and ρ_{12} are the population parameters.

Morgan tested the hypothesis that the variances of the two variables are equal, i.e.,

H_1: $$\sigma_1 = \sigma_2 .$$

[1] See bibliography at the end of the paper.

Other hypotheses that will be considered in the present paper are as follows:

H_2: Assuming $\sigma_1 = \sigma_2$; to test $\rho_{12} = \rho_0$.
H_3: Assuming $\sigma_1 = \sigma_2$; to test $\xi_1 = \xi_2$.
H_4: To test simultaneously $\sigma_1 = \sigma_2$, $\rho_{12} = \rho_0$.
H_5: To test simultaneously $\sigma_1 = \sigma_2$, $\xi_1 = \xi_2$.
H_6: Assuming $\sigma_1 = \sigma_2$ and $\xi_1 = \xi_2$; to test $\rho_{12} = \rho_0$.
H_7: Assuming $\sigma_1 = \sigma_2$, and $\rho_{12} = \rho_0$; to test $\xi_1 = \xi_2$.

Derivation of the criteria. Let x_{1i}, x_{2i} be the measurements of the two characters on the ith individual of the sample, then the joint elementary probability law of the two sets of n observations $E = (x_{11}, x_{12}, \cdots, x_{1n}; x_{21}, x_{22}, \cdots, x_{2n})$ is

$$p(E \mid \xi_1, \xi_2, \sigma_1, \sigma_2, \rho_{12}) = \left(\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}}\right)^n$$

$$\cdot \exp\left\{-\frac{1}{2(1-\rho_{12}^2)}\sum_{i=1}^{n}\left[\left(\frac{x_{1i}-\xi_1}{\sigma_2}\right)^2 - 2\rho_{12}\left(\frac{x_{1i}-\xi_1}{\sigma_1}\right)\left(\frac{x_{2i}-\xi_2}{\sigma_1}\right) + \left(\frac{x_{2i}-\xi_2}{\sigma_2}\right)^2\right]\right\}. \tag{2}$$

It will be convenient to denote by A, B, C, D, the following conditions of the population from which the sample is supposed to be drawn.

(A) that stated in equation (1).
(B) that stated in the equation for H_1, namely
$\sigma_1 = \sigma_2 = \sigma$($\sigma$ being unspecified).
(C) $\xi_1 = \xi_2 = \xi$(ξ being unspecified).
(D) $\rho_{12} = \rho_0$.

Neyman and Pearson's method affords a simple rule for obtaining appropriate test criteria once two sets of conditions have been defined. These are

(a) the conditions which can be assumed to be satisfied in any case, and
(b) the conditions which are satisfied if the hypothesis to be tested is true.

The conditions (a) define a class Ω of admissible populations, and the conditions (b) define a sub-class ω of Ω to which the population must belong if the hypothesis tested be true.

The maximum value of $p(E \mid \xi_1, \xi_2, \sigma_1, \sigma_2, \rho_{12})$ when the parameters vary in such a way that the population sampled always belongs to Ω, is called $p(\Omega \text{ max.})$. The maximum value when the population is restricted to ω is called $p(\omega \text{ max.})$. The likelihood ratio for testing the hypothesis specifying the subset ω has been defined to be

$$\lambda = \frac{p\ (\omega \text{ max.})}{p\ (\Omega \text{ max.})}. \tag{3}$$

It will be seen that $1 \leq \lambda \leq 0$. By referring λ, or a monotonic function of λ, to its sampling distribution when the hypothesis tested is true, we obtain a scale on which to assess our judgment of the truth of the hypothesis tested.

For each of the hypotheses H_1 to H_7, λ of (3) can be found. However, we shall use a more convenient criterion.

$$L = \lambda^{2/n} \tag{4}$$

which is a monotonic function of λ.

Thus the respective test criteria are found to be:

For H_1:

$$L_1 = \frac{4s_1^2 s_2^2(1 - r_{12}^2)}{(s_1^2 + s_2^2)^2(1 - R_1^2)} \tag{5}$$

where $R_1 = \dfrac{2r_{12}s_1s_2}{s_1^2 + s_2^2}$ is the estimate of ρ_{12} when σ_1 and σ_2 are assumed to be equal.

For H_2:

$$L_2 = \frac{(1 - \rho_0^2)(1 - R_1^2)}{(1 - \rho_0 R_1)^2}. \tag{6}$$

For H_3:

$$L_3 = 1 \Big/ \left\{1 + \frac{(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2 - 2r_{12}s_1s_2}\right\}. \tag{7}$$

For H_4:

$$L_4 = \frac{4(1 - \rho_0^2)s_1^2 s_2^2(1 - r^2)}{(s_1^2 + s_2^2)^2(1 - \rho_0 R_1)^2} = L_1 \times L_2. \tag{8}$$

For H_5:

$$L_5 = \frac{4s_1^2 s_2^2(1 - r_{12}^2)}{\{s_1^2 + s_2^2 + \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2\}(1 - R_2^2)} = L_1 \times L_2. \tag{9}$$

For H_6:

$$L_6 = \frac{(1 - \rho_0^2)(1 - R_2^2)}{(1 - \rho^0 R_2)^2} \tag{10}$$

where $R_2 = \dfrac{2r_{12}s_1s_2 - \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2 + \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2}$ is the estimate of ρ_{12} when both the σ's and the ξ's are assumed to be equal.

For H_7:

$$L_7 = 1 \Big/ \left\{1 + \frac{(1 + \rho_0)(\bar{x}_1 - \bar{x}_2)^2}{2(s_1^2 - 2\rho_0 r_{12}s_1s_2 + s_2^2)}\right\}^2. \tag{11}$$

The different hypotheses are also given in Table V, at the end of this paper,

together with the conditions defining sets of Ω and ω, and the appropriate likelihood criteria.

To complete the solution we must find the distributions of L or some monotonic function of L in each case when the hypothesis tested is true, in order to assess the significance of an observed value of L.

3. The distributions of the criteria. In order to simplify the problem of finding the distributions of the criteria, consider the following transformation:

$$x_{1i} = (X_i - Y_i)/\sqrt{2}$$
$$x_{2i} = (X_i + Y_i)/\sqrt{2}. \tag{12}$$

It is clear that in view of (1) X and Y will be two normally correlated variables. We shall denote this property by A' corresponding to A. The conditions B', C', D' corresponding to B, C, D respectively are as follows:

B': $\rho_{XY} = 0,$

C': $\xi_X = 0,$

D': $\sigma_Y^2 = \gamma_0 \sigma_X^2$ (when $\rho_{XY} = 0$)

where

$$\gamma_0 = \frac{1 + \rho_0}{1 - \rho_0}. \tag{13}$$

Thus we have the equivalent hypotheses H_1', $H_2' \cdots H_7'$ corresponding to H_1, H_2, $\cdots H_7$. The likelihood ratios L_1', $L_2' \cdots L_7'$ may be determined in the same way as before, and, in view of the transformation (12), it will be seen that they are equal to L_1, $L_2 \cdots L_7$ respectively.

The tests of the hypotheses H_1', H_2', H_3' are now seen to be well known.

The test of $H_1' : \rho_{XY} = 0$ is the test for significance of a correlation coefficient, and the criterion L_1 becomes

$$L_1 = \lambda_{H_1}^{2/n} = 1 - r_{XY}^2. \tag{14}$$

This test has been dealt with by Morgan [6] and Pitman [15], and has been referred to above.

The test of $H_2' : \sigma_Y^2/\sigma_X^2 = \gamma_0$ *when* $\rho_{XY} = 0$ can be treated as an extension of Fisher's z-test [5], since γ_0 is specified. If we write

$$u = \frac{S_Y^2}{S_X^2} = \frac{1 + R_1}{1 - R_1} = \frac{s_1^2 + s_2^2 + 2r_{12}s_1s_2}{s_1^2 + s_2^2 - 2r_{12}s_1s_2} \tag{15}$$

the test criterion L_2 of (6) may be written

$$L_2 = \frac{4u}{\gamma_0(1 + u/\gamma_0)^2}. \tag{16}$$

It is well known that if H_2' is true, then

$$(17)\qquad p(u) = \frac{1}{\gamma_0 B[\frac{1}{2}(n-1), \frac{1}{2}(n-1)]}\left(\frac{u}{\gamma_0}\right)^{\frac{1}{2}(n-3)}\left(1+\frac{u}{\gamma_0}\right)^{-(n-1)}$$

and the test appropriate to H_2' and therefore of H_2 is the associated z-test ($z = \frac{1}{2}\log u/\gamma_0$) with degrees of freedom $f_1 = f_2 = n - 1$. It may be easily shown that the two values of u cutting off equal tail areas from the distribution $p(u)$ will correspond to a single value of L_2.

The test of H_3' : $\xi_X = 0$ *when* $\rho_{XY} = 0$ is in the form of "Student's" t test. If we write

$$(18)\qquad \frac{t^2}{n-1} = \frac{\bar{X}^2}{s_X^2} = \frac{(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2 - 2r_{12}s_1s_2}$$

it follows that the test criterion L_3 of (12) may be written

$$(19)\qquad L_3 = 1\Big/\left(1 + \frac{t^2}{n-1}\right).$$

But it is well known that if $\xi_X = 0$, then

$$(20)\qquad p(t) = \frac{1}{\sqrt{n-1}B[\frac{1}{2}, \frac{1}{2}(n-1)]}\left(1 + \frac{t^2}{n-1}\right)^{-\frac{1}{2}n}.$$

The 5% or 1% points of significance of t may be obtained from Fisher's t-table [5] with degrees of freedom $f = n - 1$.

The tests of H_4 *and* H_5. We infer from (14), (16) and (19) that L_1 is a function of r_{XY}, L_2 a function of S_Y and S_Y, and L_3 a function of X and S_X. It is clear that if r_{XY} is distributed independently of S_X and S_Y, then L_1 and L_2 are independent, i.e.,

$$(21)\qquad p(L_1, L_2) = p(L_1)p(L_2)$$

and that if r_{XY} is distributed independently of X and S_X, then L_1 and L_3 are independent, i.e.,

$$(22)\qquad p(L_1, L_3) = p(L_1)p(L_3).$$

It is known that X, Y are independent of S_X, S_Y, r_{XY}; and in addition that r_{XY} is distributed independently of S_X, S_Y if $\rho_{XY} = 0$. Therefore, if H_1' is true, then the relations (21) and (22) hold. Hence, knowing $p(L_1)$ and $p(L_2)$, a very simple transformation and integration gives $p(L_4)$. Similarly, the distribution of L_5 may be readily derived from those of L_1 and L_3.

But from the distribution of r_{XY} when $\rho_{XY} = 0$, by transformation (14), the distribution of L_1 assuming H_1' true is found to be

$$(23)\qquad p(L_1) = \frac{1}{B[\frac{1}{2}(n-2), \frac{1}{2}]} L_1^{\frac{1}{2}(n-4)}(1 - L_1)^{-\frac{1}{2}}.$$

If H_2' is true, from (17), by transformation (16) we have

(24) $$p(L_2) = \frac{1}{B[\frac{1}{2}(n-1), \frac{1}{2}]} L_2^{\frac{1}{2}(n-3)}(1 - L_2)^{-\frac{1}{2}}.$$

Again, if H_3' is true, from (20), by transformation (19), we have

(25) $$p(L_3) = \frac{1}{B[\frac{1}{2}(n-1), \frac{1}{2}]} L_3^{\frac{1}{2}(n-3)}(1 - L_3)^{-\frac{1}{2}}$$

which is the same as the distribution of L_2. Therefore by comparing (21) and (22) we see that the distribution of L_5 when H_5' is true will be exactly the same as that of L_4 when H_4' is true. We shall therefore confine ourselves to the problem of obtaining the distribution of L_4 from those of L_1 and L_2.

Now

(26) $$p(L_1, L_2) = \frac{1}{B[\frac{1}{2}(n-2), \frac{1}{2}]B[\frac{1}{2}(n-1), \frac{1}{2}]} L_1^{\frac{1}{2}(n-4)}(1 - L_1)^{-\frac{1}{2}} L_2^{\frac{1}{2}(n-3)}(1 - L_2)^{-\frac{1}{2}}.$$

Applying the transformation

(27) $$\begin{aligned} L_4 &= L_1 L_2 \\ Z &= L_2 \end{aligned}$$

and integrating with respect to Z from 0 to 1, we obtain

(28) $$p(L_4) = \tfrac{1}{2}(n-2)L_4^{\frac{1}{2}(n-4)}, \qquad 0 \leq L_4 \leq 1.$$

Thus we can construct the values of L_4 at the 5% and 1% levels for different values of n as given in Table I.

TABLE I

5% and 1% values of L_4 (or L_5)

n	5%	1%
5	.1357	.0464
6	.2509	.1000
7	.3017	.1585
8	.3684	.2154
9	.4249	.2683
10	.4729	.3162
12	.5493	.3981
15	.6307	.4924
20	.7169	.5995
24	.7616	.6579
30	.8074	.7197
40	.8541	.7848
60	.9019	.8532
120	.9505	.9249
∞	1.0000	1.0000

The test of H_6. In the case of testing $H_6'(\sigma_Y^2 = \gamma_0\sigma_X^2)$, assuming ρ_{XY} and ρ_X each to be zero, the likelihood estimate of σ_X^2 becomes $\Sigma X^2/n$ or $S_X^2 + \bar{X}^2$. The distribution of this quantity is the same as that of S_X^2 but with degrees of freedom n instead of $n - 1$. Therefore, by analogy with the previous result (17) used in testing H_2, if we write

$$(29) \qquad v = \frac{nS_Y^2}{\Sigma X^2} = \frac{S_Y^2}{S_X^2 + \bar{X}^2} = \frac{1 + R_2}{1 - R_2}$$

then the likelihood criterion of H_6 becomes

$$(30) \qquad L_6 = \frac{4v}{\gamma_0\left(1 + \dfrac{v}{\gamma_0}\right)^2}$$

and

$$(31) \qquad p\left(v \;\middle|\; \frac{\sigma_Y^2}{\sigma_X^2} = \gamma_0\right) = \frac{1}{\gamma_0 B[\frac{1}{2}(n-1), \frac{1}{2}n]}\left(\frac{v}{\gamma_0}\right)^{\frac{1}{2}(n-3)}\left(1 + \frac{v}{\gamma_0}\right)^{-(n-\frac{1}{2})}.$$

Hence the test appropriate to H_6 is the associated z-test $z = \frac{1}{2}\log\left\{\dfrac{v}{\gamma_0}\Big/\dfrac{n-1}{n}\right\}$ with $f_1 = n - 1, f_2 = n$. We can use the z-table as before.

The test of H_7. Here we test whether $\xi_X = 0$. It may be seen that L_7 is a function of $\bar{X}^2/(S_Y^2 + \gamma_0 S_X^2)$. Further, if we assume that $\rho_{XY} = 0$ and also that $\sigma_Y^2 = \gamma_0 S_X^2$, then it will follow that $\Sigma(X - \bar{X})^2$ and $\dfrac{1}{\gamma_0}\Sigma(Y - \bar{Y})^2$ are each distributed independently as $\chi^2\sigma_X^2$ with $n - 1$ degrees of freedom; and hence their sum is distributed as $\chi^2\sigma_X^2$ with $2n - 2$ degrees of freedom. Also if $\xi_X = 0$ (and H_7' is true) X will be distributed normally about zero with standard error $\sigma_{X/\sqrt{n}}$. Hence we may write

$$(32) \qquad L_7 = 1\Big/\left\{1 + \frac{t^2}{2n-2}\right\}^2$$

where

$$(33) \qquad t^2 = \bar{X}\Big/\sqrt{\frac{\Sigma(X-\bar{X})^2 + \Sigma(Y-\bar{Y})^2/\gamma_0}{n(2n-2)}}$$

and is distributed in accordance with "Student's" distribution with $2n - 2$ degrees of freedom,

$$(34) \qquad p(t_2) = \frac{1}{\sqrt{2n-2}B[\frac{1}{2}, \frac{1}{2}(2n-2)]}\left(1 + \frac{t^2}{2n-2}\right)^{-\frac{1}{2}(2n-1)}.$$

In terms of original variables

$$(35) \qquad \frac{t_2^2}{2n-2} = \frac{\gamma_0\bar{X}^2}{\gamma_0 S_X^2 + S_Y^2} = \frac{(1+\rho_0)(\bar{x}_1 - \bar{x}_2)^2}{2(s_1^2 - 2\rho_0 r_{12}s_1 s_2 + s_2^2)}.$$

4. Comparison of the R_1-test and R_2-test with the r_{12}-test in cases where H_2 and H_6 are true respectively. It will be noted that in the preceding discussion we have been concerned with three different tests of the hypothesis that ρ_{12} has some specified value ρ_0. When there is no information available regarding the means and standard deviations of x_1 and x_2, the test is based on the sampling distribution of the ordinary product-moment coefficient r_{12}. If it may be assumed that $\sigma_1 = \sigma_2$, then we have the estimate

$$R_1 = \frac{2r_{12}s_1s_2}{s_1^2 + s_2^2}.$$

If besides $\sigma_1 = \sigma_2$, it may also be assumed that $\xi_1 = \xi_2$, then we have the estimate

$$R_2 = \frac{2r_{12}s_1s_2 - \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2 + \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2}.$$

From the point of view of testing hypotheses, all these criteria r_{12}, R_1, R_2 follow from the application of the likelihood ratio method. It will be noted that if $\sigma_1 = \sigma_2$, either the r_{12} or the R_1 test may be used. But, insofar as the likelihood principle is accepted, the latter should be regarded as the "better" test. Again, if $\sigma_1 = \sigma_2$ and $\xi_1 = \xi_2$, all three tests may be used, but that based on R_2 will be the "best". A question of interest is to investigate just what is meant by the "better" or the "best" test. We may ask how far the improvements are sufficient to justify the use of the R_1 and R_2 tests in place of the more generally used r_{12} test. One method of comparison is to examine what Neyman and Pearson [12] have termed the "power function" of the tests.

For example, when testing the hypothesis that a parameter θ has the value θ_0 in the population sampled, the power of the test criterion T with regard to the alternative hypothesis that $\theta = \theta_1 > \theta_0$ is given by the expression $\beta(\theta_1) = P\{T > T_\alpha | \theta = \theta_1\}$ where $T($ is the value of T at the level of significance α. This quantity $\beta(\theta)$ measures the chance that the test as specified will detect the fact that $\theta = \theta_0$, i.e., the chance of rejecting the hypothesis when it is not true. A test whose power function is never less than that of any other test is termed the uniformly most powerful test.

If the permissible alternative hypotheses to $\theta = \theta_0$ are both $\theta < \theta_0$ and $\theta > \theta_0$, then the power of the test T is given by the expression

$$\beta(\theta_1) = 1 - p\{T'_\alpha < T < T''_\alpha | \theta_1\}$$

where T'_α and T''_α are the values of T at both ends of the distribution at the level of the significance α. When the test is such that the power function has a minimum value α at $\theta = \theta_0$, it is said to be unbiased.

A test is termed biased if, for certain alternative hypotheses $\theta \neq \theta_0$, the chance of rejecting the hypothesis $\theta = \theta_0$ is less than the chance of rejecting this hypothesis when it is true.

In what follows it is proposed to compare the power functions of the tests based on r_{12}, R_1, and R_2 in order to obtain more complete evidence of the extent to which one is "better" than the other.

The distribution of R_1.[2] We have obtained the distribution of n when H_2' and therefore H_2 is true. We are now able to find the distribution of R_1 by applying the transformation of (15). Thus the distribution of R_1 in terms of ρ_0 is

$$p(R_1 \mid \rho_0) = \frac{(1-\rho_0^2)}{2^{n-2}B[\frac{1}{2}(n-1), \frac{1}{2}(n-1)]} \frac{(1-R_1^2)^{\frac{1}{2}(n-3)}}{(1-\rho_0 R_1)^{n-1}}. \tag{36}$$

The significance of R_1 may be assessed by the z-test, where we take

$$\begin{aligned} Z &= \frac{1}{2}\log_e \frac{u}{\gamma_0} = \frac{1}{2}\log\frac{1+R_1}{1-R_1} - \frac{1}{2}\log\frac{1+\rho_0}{1-\rho_0}. \\ &= z' - \zeta, \text{ say} \end{aligned} \tag{37}$$

with degrees of freedom $f_1 = f_2 = n - 1$. R. A. Fisher's z-table may be used in this connection.

When $\rho_{12} = 0$, the distribution simplifies to

$$\begin{aligned} p(R_1 \mid \rho_{12} = 0) &= \frac{1}{2^{n-1}B[\frac{1}{2}(n-1), \frac{1}{2}(n-1)]}(1-R_1^2)^{\frac{1}{2}(n-3)} \\ &= \frac{1}{B[\frac{1}{2}(n-1), \frac{1}{2}]}(1-R_1^2)^{\frac{1}{2}(n-3)} \end{aligned} \tag{38}$$

since 2^{2n-2} $B[\frac{1}{2}(n-1), \frac{1}{2}(n-1)]$ is equal to $B[\frac{1}{2}(n-1), \frac{1}{2}]$ by duplication formula [16, p. 240].

The distribution (38) is similar in form to that of $p(r_{12} \mid \rho_{12} = 0)$ with $n - 1$ degrees of freedom instead of $n - 2$. The significance levels of R_1 may then be obtained directly from the r-table [1] for the case $\rho_{12} = 0$, entering with degrees of freedom $n - 1$.

The distribution of R_2. The distribution of R_2 may be obtained from that of v when H_6' and therefore H_6 is true. It is

$$p(R_2 \mid \rho_{12} = \rho_0) = \frac{(1+\rho_0)^{\frac{1}{2}n}(1-\rho_0)^{\frac{1}{2}(n-1)}}{2^{n-\frac{1}{2}}B[\frac{1}{2}(n-1), \frac{1}{2}n]} \frac{(1+R_2)^{\frac{1}{2}(n-3)}(1-R_2)^{\frac{1}{2}(n-2)}}{(1-\rho_0 R_2)^{n-\frac{1}{2}}}. \tag{39}$$

This agrees with the result first obtained by R. A. Fisher [4] by a different method. The significance of R_2 may be assessed by the z-test, where we take

$$z = \frac{1}{2}\log\left(\frac{v}{\gamma_0} \Big/ \frac{n-1}{n}\right) \tag{40}$$

[2] Since finding the distribution of R_1 (36), (38) and the relation between R_1 and z' (37), my attention has been drawn to a recent paper by DeLury [2] in which the same results are obtained. Since my method of derivation is different from his, I have thought it worthwhile to retain it here.

with degrees of freedom $f_1 = n - 1, f_2 = n$. The tables for use with the z-test may be used in this connection.

When $\rho_{12} = 0$, the distribution is simplified to

$$(41) \qquad p(R_2 \mid \rho_{12} = 0) = \frac{1}{2^{n-\frac{1}{2}}B[\frac{1}{2}(n-1), \frac{1}{2}n]} (1 + R_2)^{\frac{1}{2}(n-3)}(1 - R_2)^{\frac{1}{2}(n-2)}$$

which is simply a Pearson Type I curve.

Power functions of R_1 and R_2. In order to find the power functions of R_1 and R_2 with respect to alternative hypotheses H_t to H_2, specifying $\rho_{12} = \rho_t < \rho_0$, it will be convenient to consider the incomplete beta function distributions

$$(42) \qquad p(x_1) = \frac{1}{B[\frac{1}{2}(n-1), \frac{1}{2}(n-1)]} x_1^{\frac{1}{2}(n-3)}(1 - x_1)^{\frac{1}{2}(n-3)}$$

$$(43) \qquad p(x_2) = \frac{1}{B[\frac{1}{2}(n-1), \frac{1}{2}n]} x_2^{\frac{1}{2}(n-3)}(1 - x_2)^{\frac{1}{2}(n-2)}$$

where $x_1 = \dfrac{u}{\gamma_0(1 + u/\gamma_0)}$ and $x_2 = \dfrac{v}{\gamma_0(1 + v/\gamma_0)}$. From the *Tables of the Incomplete Beta Function* [13] we can find the values of x_1 and x_2 at the significance level α, i.e.

$$(44) \qquad I_{x_1'}[\tfrac{1}{2}(n-1), \tfrac{1}{2}(n-1)] = \alpha',$$

$$(45) \qquad I_{x_2'}[\tfrac{1}{2}(n-1), \tfrac{1}{2}n] = \alpha'.$$

The values of $R_1'(\alpha)$, and of $R_2'(\alpha)$, may then be calculated from the relations

$$(46) \qquad R_1 = \frac{u-1}{u+1} = \frac{-1 + x_1 + \gamma_0 x_1}{1 - x_1 + \gamma_0 x_1},$$

$$(47) \qquad R_2 = \frac{v-1}{v+1} = \frac{-1 + x_2 + \gamma_0 x_2}{1 - x_2 + \gamma_0 x_2}.$$

The power functions of R_1 and R_2 thus found may be given as follows:

$$(48) \qquad \beta'(\rho_t \mid R_1) = P\{R_1 < R_1'(\alpha) \mid \rho_t\},$$

$$(49) \qquad \beta'(\rho_t \mid R_2) = P\{R_2 < R_2'(\alpha) \mid \rho_t\}.$$

In the same way, for any alternative hypothesis H_t specifying $\rho_{12} = \rho_t > \rho_0$, we can find the values of x_1 and x_2 at the significance level α'', at the other end of the distribution, i.e.

$$(50) \qquad 1 - I_{x_1''}[\tfrac{1}{2}(n-1), \tfrac{1}{2}(n-1)] = \alpha'',$$

$$(51) \qquad 1 - I_{x_2''}[\tfrac{1}{2}(n-1), \tfrac{1}{2}n] = \alpha''.$$

Thence the corresponding values of $R_1''(\alpha)$ and $R_2''(\alpha)$ may be obtained, and their power functions are

$$(52) \qquad \beta''(\rho_t \mid R_1) = P\{R_1 > R_1''(\alpha) \mid \rho_t\},$$

$$\beta''(\rho_t \mid R_2) = P\{R_2 > R_2''(\alpha) \mid \rho_t\}. \tag{53}$$

The power functions of R_1 and R_2 with respect to alternative hypotheses specifying $\rho_{12} = \rho_t < \rho_0$ and $> \rho_0$ may now be obtained by adding (48) and (52) or (49) and (53) or, more simply,

$$\beta(\rho_t \mid R_1) = 1 - P\{R_1'(\alpha) < R_1 < R_1''(\alpha) \mid \rho_t\}, \tag{54}$$

$$\beta(\rho_t \mid R_2) = 1 - P\{R_2'(\alpha) < R_2 < R_2''(\alpha) \mid \rho_t\} \tag{55}$$

where $R_1'(\alpha)$, $R_1''(\alpha)$; $R_2'(\alpha)$, $R_2''(\alpha)$ are the values of R_1 and R_2 at the two ends of the distribution at the significance level $\alpha = \alpha' + \alpha''$.

In view of the fact that after transformation the tests based on R_1 and R_2 are equivalent to tests regarding the equality of variances, it follows from Neyman and Pearson's work [11] regarding the uniformly most powerful test of the hypothesis that $\sigma_Y^2/\sigma_X^2 = \gamma_0$, with alternatives $\sigma_Y^2/\sigma_X^2 = \gamma_t < \gamma_0$ (or $\gamma_t > \gamma_0$), that: (1) if $\sigma_1 = \sigma_2$ and alternative to $\rho_{12} = \sigma_0$ are that $\rho_{12} = \rho_t < \rho_0$ (or, in a second case, $\rho_t > \rho_0$) the test based on R_1 is the uniformly most powerful test, i.e., it is more powerful than that based on r_{12}; and (2) if $\sigma_1 = \sigma_2$ and $\xi_1 = \xi_2$, then the test based on R_2 is the uniformly most powerful test, i.e., it is more powerful than those based on either r_{12} or R_1.

For illustration, let us take a special case, say

$$n = 10, \qquad \rho_0 = 0.6, \qquad \alpha' = \alpha'' = 0.025. \tag{a}$$

From the tables, we obtain the values

$$x_1' = .198902 \qquad x_2' = .184863$$
$$x_1'' = .801098 \qquad x_2'' = .772916$$

and by calculation the values

$$R_1'(\alpha) = -.0034 \qquad R_2'(\alpha) = -.0487$$
$$R_1''(\alpha) = .8831 \qquad R_2''(\alpha) = .8632.$$

The values of the power functions of R_1 and R_2 for specified values of ρ_t have been calculated and are given in Table II. For $\rho_t < \rho_0$, a comparison of columns 2 and 4 will show that the test based on R_2 is uniformly more powerful than that based on R_1 (or for $\rho_t > \rho_0$, a comparison of columns 3 and 5).

The unbiased test of H_2 and H_6. When however the alternatives are that $\rho_{12} = \rho_t < \rho_0$, and $\rho_t > \rho_0$, questions of bias may be introduced.

In the case of H_2, i.e. when R_1 is used, it was established by J. Neyman in his lecture courses [8], that if we test whether $\sigma_Y^2/\sigma_X^2 = \gamma_0$, where the alternatives are $\gamma_t < \gamma_0$ and $\gamma_t > \gamma_0$, and if the samples of X and Y are of equal size, then the test based on cutting off equal tail areas of the distribution of x_1 is unbiased and of the type B [7]. Therefore the same may be said of the R_1-test.

In the case of H_6, the equivalent transformed test is again whether $\sigma_Y^2/\sigma_X^2 = \gamma_0$. But the test now corresponds to that in which an estimate of σ_Y^2 is based

on $f_1 = n - 1$ degrees of freedom and an estimate of σ_X^2 on $f_2 = n$ degrees of freedom. The degrees of freedom not being equal, it is known that if equal tail areas are cut off from the sampling distribution of x_2, this test will be biased. Neyman's result [8] shows that if the lower and upper significance levels are taken at x_2' and x_2'', then the equation

$$x_2''^{f_1}(1 - x_2'')^{f_2} = x_2'^{f_1}(1 - x_2')^{f_2} \tag{56}$$

should be satisfied if the test is unbiased. Since in the present case, with the test based on equal tail area critical region, the bias will be very small, the rejection levels $R_2'(\alpha)$ and $R_2''(\alpha)$ in the numerical investigation given in Table III have been selected taking equal tail areas for simplicity.

TABLE II

Values of the power functions of R_1 and R_2 with respect to alternative hypotheses

$\rho_{12} = \rho_t < \rho_0$ or $\rho_t > \rho_0$

$(n = 10; \rho_0 = 0.6; \alpha' = \alpha'' = 0.025)$

ρ_t	$\beta'(\rho_t \mid R_1)$	$\beta''(\rho_t \mid R_1)$	$\beta'(\rho_t \mid R_2)$	$\beta''(\rho_t \mid R_2)$
−0.8	.9984			
−0.6	.9739		.9807	
−0.4	.9867		.9005	
−0.2	.7189		.7360	
0.0	.4960	.0002	.5093	.0001
0.2	.2744	.0008	.2809	.0006
0.3	.1825	.0018	.1860	.0015
0.4	.1106	.0042	.1111	.0037
0.5	.0576	.0099	.0580	.0093
0.6	.025	.025	.025	.025
0.7	.0081	.0678	.0080	.0720
0.8	.0015	.1995	.0015	.2150
0.9	.0001	.5950	.0001	.6289
0.95		.8979		.9150
0.975		.9866		.9897

If we now take a special case, similar to (a) above, but taking equal tail areas, so that

$$n = 10 \qquad \rho = 0.6$$

$$\alpha = 0.5 \qquad (\alpha' = \alpha'' = \tfrac{1}{2}\alpha)$$

we can obtain the values of x's and of R's as before.

The values of the power functions of R_1 and R_2 for specified values of ρ_t are given in columns 3 and 4 of Table III. These values are equivalent to the sums of the corresponding values in Table II. The values of the power functions of R_1 and R_2 for the following additional cases are also given in Table III:

(b) $n = 10 \quad \rho_0 = 0.8 \quad \alpha = 0.05$

(c) $n = 20 \quad \rho_0 = 0.6 \quad \alpha = 0.05$

(d) $n = 20 \quad \rho_0 = 0.8 \quad \alpha = 0.05.$

Comparison of the power functions. We may now deal with the question raised at the beginning of this section, namely, as to what is meant by the "better" or "best" test. We shall proceed to compare for certain special cases the power functions of the three test, all of which are applicable where it may be assumed that $\sigma_1 = \sigma_2$, $\xi_1 = \xi_2$.

In the first place it will be noted that the power function of the test based on equal tail areas of the r_{12} distribution is

$$\beta(\rho_t \mid r_{12}) = 1 - p\{\gamma'_{12}(\alpha) < \gamma_{12} < \gamma''_{12}(\alpha) \mid \rho_t\} \tag{57}$$

where

$$\begin{aligned} P\{r_{12} < r'_{12}(\alpha) \mid \rho_0\} &= \int_{-1}^{r'_{12}(\alpha)} p(r_{12} \mid \rho_{12} = \rho_0)\, dr_{12} = \tfrac{1}{2}\alpha \\ P\{r_{12} > r'_{12}(\alpha) \mid \rho_0\} &= \int_{r''_{12}(\alpha)}^{1} p(r_{12} \mid \rho_{12} = \rho_0)\, dr_{12} = \tfrac{1}{2}\alpha \end{aligned} \tag{58}$$

and

$$p(r_{12} \mid \rho_{12} = \rho_0) = \frac{(1 - \rho_0^2)^{\frac{1}{2}(n-1)}}{\pi\Gamma[\frac{1}{2}(n-1)]}(1 - r_{12}^2)^{\frac{1}{2}(n-4)}\left(\frac{\partial}{\partial r_{12}}\right)^{n-2} \frac{\cos^{-1}(-\rho_0 r_{12})}{\sqrt{(1 - \rho_0^2 r_{12}^2)}}. \tag{59}$$

The probability that r_{12} is less than some specified value may be obtained from *Tables of the Correlation Coefficient* (F. N. David, [1]), or, where these are not sufficiently detailed, by using R. A. Fisher's z'-transformation for r_{12} [4].

The cases considered are (a), (b), (c), (d) as defined above. The power functions of the three different tests (all based upon the equal tail areas of their distributions) are given in Table III. The figures for r_{12} in the brackets are those obtained by the z'-transformation approximation.

An examination of Tables II and III brings out the following points:

(1) For reasons given above, the R_2 test based on equal tail area critical regions is very slightly biased; the amount of this bias for the case $n = 10$, $\rho_0 = 0.6$, $\alpha = 0.05$ is shown in Table IV. This shows that the power of the R_2 test is less than 0.05 in the fifth or sixth decimal places for $0.59 < \rho_t < 0.60$. As a result this test is very slightly less powerful than the other two tests for alternatives with ρ_t slightly less than ρ_0. The effect is, however, of little importance.

(2) Except in this short range of ρ_t, we find that

$$\beta(\rho_t \mid R_2) \geq \beta(\rho_t \mid R_1) \geq \beta(\rho_t \mid r_{12}).$$

TABLE III

Comparison of the power functions of r_{12}, R_1, and R_2 tests with respect to alternative hypotheses

ρ_t	$n = 10$ $\rho_0 = 0.6$			$n = 10$ $\rho_0 = 0.8$			$n = 20$ $\rho_0 = 0.6$			$n = 20$ $\rho_0 = 0.8$		
	$\beta(\rho_t \mid r_{12})$	$\beta(\rho_t \mid R_1)$	$\beta(\rho_t \mid R_2)$	$\beta(\rho_t \mid r_{12})$	$\beta(\rho_t \mid R_1)$	$\beta(\rho_t \mid R_2)$	$\beta(\rho_t \mid r_{12})$	$\beta(\rho_t \mid R_1)$	$\beta(\rho_t \mid R_2)$	$\beta(\rho_t \mid r_{12})$	$\beta(\rho_t \mid R_1)$	$\beta(\rho_t \mid R_2)$
−0.6	.9739	.9739	.9807									
−0.4	.8865	.8867	.9005	.9887	.9891	.9921	.9965	.9967	.9973			
−0.2	.7186	.7189	.7360	.9557	.9569	.9650	.9648	.9663	.9698			
0.0	.4960	.4962	.5094	.9742	.8766	.8909	.8328	.8369	.8449	.9952	.9959	.9966
0.2	.2753	.2752	.2815	.7158	.7189	.7360	.5412	.5456	.5534	.9624	.9663	.9698
0.4	.1142	.1148	.1148	.4727	.4750	.4877	.2026	.2036	.2061	.8062	.8170	.8254
0.5	.0679	.0675	.0673	.3330	.3345	.3427	.0915	.0917	.0922	.6309	.6432	.6520
0.6	.0500	.0500	.0500	.2005	.2010	.2047	.0500	.0500	.0500	.3920	.4011	.4085
0.7	.0735	.0759	.0800	.0969	.0965	.0971	.1096	.1119	.1147	.1589	.1617	.1635
0.8	.1890	.2010	.2165	.0500	.0500	.0500	.3886	.4010	.4134	.0500	.0500	.0500
0.9	.5656	.5951	.6290	.1466	.1771	.1904	.9034	.9106	.9181	.3272	.3493	.3604
	(.5569)			(.1454)			(.9004)			(.3270)		
0.95	(.8709)	.8979	.9150	(.4689)	.5426	.5763	(.9974)	.9974	.9978	(.8547)	.5871	.8844
0.975	(.9822)	.9866	.9897	(.8134)	.8692	.8896				(.9944)	.9947	.9960
0.99				(.9845)	.9908	.9938						
Levels	−.0039	−.0034	−.0487	.4004	.3817	.3423	.2289	.2253	.2041	.5671	.5613	.5459
	.8998	.8831	.8632	.9574	.9463	.9368	.8300	.8201	.8084	.9222	.9158	.9101

That is to say, the power function of the R_2 test never lies below those of the R_1 and r_{12} tests, and that of the R_1 test never lies below that of the r_{12} test.

(3) The gain in sensitivity as measured by the chance that the test will detect that $\rho_t \neq \rho_0$ is, however, very small. Further, R_1 may only be used if it is known that $\sigma_1 = \sigma_2$ and R_2 if it is known in addition that $\xi_1 = \xi_2$. It will only be in rather special problems that the statistician can feel confident that such assumptions are justified. We will therefore probably prefer the test based on the ordinary product moment correlation coefficient r_{12}, since the slight loss in power will be felt to be outweighed by the gain in simplicity. It is, however, only after an objective comparison of the consequences of applying the three tests that a definite opinion on these points can be reached.

TABLE IV

ρ_t	$\beta'(\rho_t \vert R_2)$	$\beta''(\rho_t \vert R_2)$	$\beta(\rho_t \vert R_2)$
0.5	.0580	.0093	.0673
0.590	.0274235	.0225806	.0500041
0.591	.0271778	.0228190	.0499968
0.592	.0269359	.0230578	.0499937
0.593	.0266934	.0232976	.0499910
0.594	.0264515	.0235337	.0499852
0.595	.0262096	.0237798	.0499894
0.596	.0259677	.0240222	.0499899
0.597	.0257257	.0242651	.0499908
0.598	.0254838	.0245107	.0499945
0.599	.0252419	.0247540	.0499959
0.6	.025	.025	.05

5. Summary. Various hypotheses relating to a population of two normal correlated variates have been considered and the appropriate test criteria for each hypothesis have been derived by the likelihood ratio method. The distributions of the likelihood ratio criteria or of monotonic functions of them have been obtained with the aid of transformation (14). References have been given to tables from which significance levels for use in conjunction with the tests may be obtained; a new table of significance levels for the tests of H_4 and H_5 was given.

The power functions of r_{12}, R_1 and R_2 have been compared; from these power functions it was concluded that R_1 and R_2 are suitable respectively for testing the hypothesis when $\sigma_1 = \sigma_2$ and when, in addition, $\xi_1 = \xi_2$.

In conclusion, I should like to express my indebtedness to Professor E. S. Pearson for continued advice and help in the preparation of this paper, to Dr. A. Wald and Professor S. S. Wilks for valuable suggestions.

TABLE V

Conditions defining Ω and ω together with the likelihood criteria appropriate for testing the hypotheses H_i

(1) Hypotheses H_i	(2) Initial Assumptions (Apart from Normality)	(3) To be tested	(4) Conditions defining (a) Ω	(5) Conditions defining (b) ω	(6) Criteria $L_i = \lambda_{H_i}^{2/n}$
H_1	None	$\sigma_1 = \sigma_2$	A	A, B	$\dfrac{4s_1^2 s_2^2(1 - r_{12}^2)}{\{(s_1^2 + s_2^2)^2 - 4r_{12}^2 s_1^2 s_2^2\}}$
H_2	$\sigma_1 = \sigma_2$	$\rho_{12} = \rho_0$	A, B	A, B, D	$\dfrac{(1 - \rho_0^2)(1 - R_1^2)}{(1 - \rho_0 R_1)^2}$ [1]
H_3	$\sigma_1 = \sigma_2$	$\xi_1 = \xi_2$	A, B	A, B, C	$1 \Big/ \left\{1 + \dfrac{(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2 - 2r_{12}s_1 s_2}\right\}$
H_4	None	$\sigma_1 = \sigma_2$ $\rho_{12} = \rho_0$	A,	A, B, D	$\dfrac{4s_1^2 s_2^2(1 - \rho_0^2)(1 - r_{12}^2)}{(s_1^2 + s_2^2)(1 - \rho_0 R_1)^2}$
H_5	None	$\sigma_1 = \sigma_2$ $\xi_1 = \xi_2$	A,	A, B, C	$\dfrac{4s_1^2 s_2^2(1 - r_{12}^2)}{\{s_1^2 + s_2^2 + \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2\}(1 - R_2^2)}$ [2]
H_6	$\sigma_1 = \sigma_2$ $\xi_1 = \xi_2$	$\rho_{12} = \rho_0$	A, B, C	A, B, C, D	$\dfrac{(1 - \rho_0^2)(1 - R_2^2)}{(1 - \rho_0 R_2)^2}$
H_7	$\sigma_1 = \sigma_2$ $\rho_{12} = \rho_0$	$\xi_1 = \xi_2$	A, B, D	A, B, D, C	$1 \Big/ \left\{1 + \dfrac{(1 + \rho_0)(\bar{x}_1 - \bar{x}_2)^2}{2(s_1^2 - 2\rho_0 r_{12} s_1 s_2 + s_2^2)}\right\}^2$

[1] $R_1 = \dfrac{2r_{12}s_1 s_2}{s_1^2 + s_2^2}$ [2] $R_2 = \dfrac{2r_{12}s_1 s_2 - \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2 + \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2}$

REFERENCES

[1] F. N. David, *Tables of the Correlation Coefficient*, London: Biometrika Office, 1938.
[2] D. B. DeLury, *Ann. Math. Stat.* Vol. 9 (1938) p. 149.
[3] D. J. Finney, *Biometrika*, Vol. 30 (1938), p. 190.
[4] R. A. Fisher, *Metron*, Vol. 1 (1921), p. 3.
[5] R. A. Fisher, *Statistical Methods for Research Workers*, 7th ed. Edinburgh: Oliver Boyd, 1938.
[6] W. A. Morgan, *Biometrika*, Vol. 31 (1939), p. 13.
[7] J. Neyman, *Bull. Soc. Math. France*, Vol. 63 (1935), p. 246.
[8] J. Neyman, Lectures delivered in London, 1937-8, (Unpublished).
[9] J. Neyman and E. S. Pearson, *Biometrika*, Vol. 20A (1928), p. 175.
[10] J. Neyman and E. S. Pearson, *Bull. Acad. Polonaise Sci. Lettres*, A (1931), p. 460.
[11] J. Neyman and E. S. Pearson, *Phil. Trans.*, Ser. A, Vol. 236 (1933), p. 289.
[12] J. Neyman and E. S. Pearson, *Stat. Res. Mem.* Vol. 1 (1936), p. 1.
[13] K. Pearson (Editor), *Tables of the Incomplete Beta-Function*, London: Biometrika Office.
[14] E. S. Pearson and J. Neyman, *Bull. Acad. Polonaise Sci. Lettres*, A (1930), p. 73.
[15] J. C. Pitman, *Biometrika*, Vol. 31 (1939), p. 9.
[16] E. T. Whittaker and G. N. Watson, *Modern Analysis*, 4th Edition (1927).

University College,
London, England.

ON A LEAST SQUARES ADJUSTMENT OF A SAMPLED FREQUENCY TABLE WHEN THE EXPECTED MARGINAL TOTALS ARE KNOWN

By W. Edwards Deming and Frederick F. Stephan

1. Introduction. There are situations in sampling wherein the data furnished by the sample must be adjusted for consistency with data obtained from other sources or with deductions from established theory. For example, in the 1940 census of population a problem of adjustment arises from the fact that although there will be a complete count of certain characters for the individuals in the population, considerations of efficiency will limit to a sample many of the cross-tabulations (joint distributions) of these characters. The tabulations of the sample will be used to estimate the results that would have been obtained from cross-tabulations of the entire population.[1] The situation is shown in Fig. 1 in parallel tables for the universe and for the sample. For the universe the marginal totals $N_{i.}$ and $N_{.j}$ are known, but not the cell frequencies N_{ij}; for the sample, however, tabulation gives both the cell frequencies n_{ij} and the marginal totals $n_{i.}$ and $n_{.j}$.

In estimating any cell frequency of the universe, such as N_{ij}, three possibilities present themselves; from the sample one may make an estimate from the ith row alone, another from the jth column alone, and still another from the over-all ratio n_{ij}/n: specifically, the three estimates would be $n_{ij}N_{i.}/n_{i.}$, $n_{ij}N_{.j}/n_{.j}$, and $n_{ij}N/n$. As a result of sampling errors these will not be identical except by accident, and though any of them by itself may be considered accurate enough, still, if the whole $r \times s$ table of universe cell frequencies were so estimated, the marginal totals would not come out right. In this paper we present a rapid method of adjustment, which in effect combines all three of the estimates just mentioned, and at the same time enforces agreement with the marginal totals. The method is extended to varying degrees of cross-tabulation in three dimensions.

In any problem of adjustment where the conditions are intricate it is necessary to have a method that is straight-forward and self-checking; this becomes imperative when we realize that in the three-dimensional Case VII of the problem now at hand (*vide infra*), any adjustment in one cell must be balanced by adjustments in at least seven others. The method of least squares is one possible procedure for effecting an adjustment and at the same time enforcing certain conditions among the marginal totals. It is essentially a scheme for

[1] Examples will occur in the 1940 census publications. Further discussion of this problem and of the sampling procedure is given by the authors in "The sampling procedure of the 1940 population census," *Jour. Am. Stat. Assn.*, Vol. 35 (1940), pp. 615-630.

arriving at a set of calculated or adjusted observations that will satisfy the conditions of the problem, and at the same time minimize the sum of the weighted squares of the residuals, symbolized as

$$S = \Sigma\, w(n_c - n_0)^2 \tag{1}$$

n_c and n_0 being the calculated and observed numbers in a cell, and $n_c - n_0$ the corresponding residual. It is the nature of the conditions imposed on the adjusted values that distinguishes one type of problem from another. Least squares has the practical advantage of uniqueness, once the weights of the observations have been assigned, and it possesses the theoretical dignity of giving one kind of "best" estimates under ideal conditions of sampling. For our present purpose we shall minimize sums of the form

$$S = \Sigma(m_i - n_i)^2/n_i \tag{2}$$

n_i being the observed frequency in the ith cell, and m_i the calculated or adjusted frequency therein. The conditions among the m_i will arise from the fact that the marginal totals, after adjustment, must agree with their expected values, namely, the deflated marginal totals of the universe (for example, $m_{i.}$ and $m_{.j}$ as defined in eqs. (6) and (7)).

By definition, weight and variance are inversely proportional, hence the principle of least squares is identical with the minimizing of chi-square. Here the variance in the ith cell is $\nu_i(1 - \nu_i/n)$, where ν_i is the expected number in that cell, and n is the total number in the sample. Now if ν_i is sufficiently well approximated by n_i, it follows that if no cell contains an appreciable fraction of the whole sample (a circumstance requiring a fair sized number of cells—perhaps 100), the variance may be taken as ν_i for every i, and the minimized S can be used as chi-square. But regardless of the number of cells, if the n_i be not too much different from one another, so that the factor $1 - \nu_i/n$ may be treated as a constant, we still get the least squares solution by minimizing S as defined in eq. (2).

2. The two dimensional problem. Suppose that the data on two characteristics (e.g. age and highest grade of school completed) are obtained for each member of a universe of N individuals, and that tabulations of the data provide either (a) one set of marginal totals $N_{1.}, N_{2.}, \cdots, N_{r.}$; or ($b$) in addition, the marginal totals $N_{.1}, N_{.2}, \cdots, N_{.s}$. The nature of the tabulations is presumed such that it is not feasible to count the numbers N_{ij} in the cells, as would be done if one character were crossed with the other. Suppose, however, that for a sample of n individuals selected in a random manner from the universe, the two characters are crossed with each other, so that we know not only all the $s + r$ marginal totals $n_{.1}, \cdots, n_{r.}$ of the sample but also the numbers n_{ij} ($i = 1, 2, \cdots, r; j = 1, 2, \cdots, s$). The problem is to estimate the unknown frequencies N_{ij} in the cells of the universe. This will be done by finding the calculated or adjusted sample frequencies m_{ij} and then inflating them by the inverse sampling ratio N/n.

For the least squares solution we seek those values of m_{ij} that minimize[2]

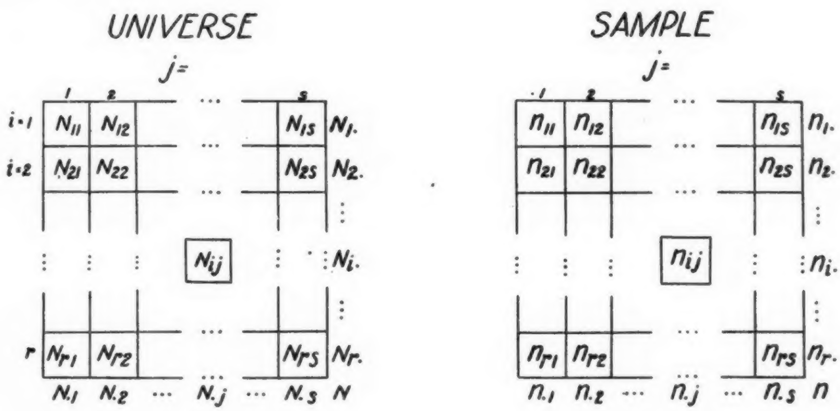
$$S = \Sigma(m_{ij} - n_{ij})^2/n_{ij} \tag{3}$$

wherein the m_{ij} are subjected to one of the following sets of conditions:

Case I: One set of marginal totals known. Assume $N_{1.}, N_{2.}, \cdots, N_{r.}$ to be known. Then we require

$$\sum_j m_{ij} = m_{i.}, \qquad i = 1, 2, \cdots, r. \tag{4}$$

These r equations constitute r conditions on the adjusted m_{ij}.



N_{ij} unknown
Marginal totals $N_{.j}$ and $N_{i.}$ known
N known

n_{ij} known
Marginal totals $n_{.j}$ and $n_{i.}$ known
n known

FIG. 1. SHOWING THE SYSTEM OF NOTATION FOR THE CELL FREQUENCIES AND MARGINAL TOTALS OF THE UNIVERSE AND THE SAMPLE IN THE TWO DIMENSIONAL PROBLEM

Case II: Both sets of marginal totals known. Here the adjusted cell frequencies must satisfy not only condition (4) but also

$$\sum_i m_{ij} = m_{.j} \qquad j = 1, 2, \cdots, s - 1 \tag{5}$$

there being now a total of $r + s - 1$ conditions. In both cases,

$$m_{i.} = N_{i.}n/N, \tag{6}$$

$$m_{.j} = N_{.j}n/N. \tag{7}$$

In other words, $m_{i.}$ and $m_{.j}$ are the deflated marginal totals, i.e., $N_{i.}$ and $N_{.j}$ divided by the actual sampling ratio N/n. The $m_{i.}$ and $m_{.j}$ are not independent, for

[2] The sign $\sum$ will denote summation over all possible cells, unless otherwise noted. $\sum_i$ will denote summation over all values of i, and similarly for an inferior j or k. The dot, as in $n_{.j}$, will signify the result of summing the n_{ij} over all values of i in the jth column.

$$N_{.1} + N_{.2} + \cdots + N_{.s} = N_{1.} + N_{2.} + \cdots + N_{r.} = N. \tag{8}$$

It is for this reason that if i runs through all r values in eq. (4), then j can run through only $s - 1$ in eq. (5). A similar equation also exists for the marginal totals of the sample, namely,

$$n_{.1} + n_{.2} + \cdots + n_{.s} = n_{1.} + n_{2.} + \cdots + n_{r.} = n. \tag{9}$$

Solution of the two dimensional Case I. Assuming that the adjusted values of the m_{ij} have been found, let each take on a small variation δm_{ij}; then the differentials of eqs. (3) and (4) show that

$$\tfrac{1}{2}\delta S = \Sigma\{(m_{ij} - n_{ij})/n_{ij}\}\delta m_{ij} = 0 \qquad \text{(one equation)}, \tag{10}$$

$$\sum_j \delta m_{ij} = 0, \qquad i = 1, 2, \cdots, r \qquad (r \text{ equations}). \tag{11}$$

Multiply now eq. (11i) by the arbitrary Lagrange multiplier $-\lambda_{i.}$, and add eqs. (10) and (11) to obtain

$$\Sigma\{(m_{ij} - n_{ij})/n_{ij} - \lambda_{i.}\}\delta m_{ij} = 0. \qquad \text{(one equation)}. \tag{12}$$

By the usual argument, one may now set each brace equal to zero, recognizing that the r Lagrange multipliers are then no longer arbitrary but must satisfy the relation

$$m_{ij} = n_{ij}(1 + \lambda_{i.}). \tag{13}$$

The adjusted frequencies m_{ij} can be computed at once as soon as the $\lambda_{i.}$ are found. To evaluate them one may rewrite the conditions (4) using the right-hand member of (13) for m_{ij}, obtaining

$$m_{i.} = n_{i.}(1 + \lambda_{i.}). \tag{14}$$

Another way to arrive at this same relation is to sum each member of eq. (13) in the ith row. However obtained $\lambda_{i.}$ is now known, since $m_{i.}$ and $n_{i.}$ are known, and in fact eq. (13) now gives

$$m_{ij} = n_{ij}(m_{i.}/n_{i.}). \tag{15}$$

The adjustment is thus a simple proportionate one by rows, the cells in any one row all being raised or lowered by the proportionate adjustment in the row total. Case I thus amounts to r independent one dimensional proportionate adjustments, one for each row, and any one or all may be carried out, as desired. This result can be obtained by a simpler approach but is presented in this way for consistency with later cases.

The minimized sum of squares may be computed directly, or from the row totals by seeing that

$$S = \sum_i (m_{i.} - n_{i.})^2/n_{i.}. \tag{16}$$

The term $(m_{i.} - n_{i.})^2/n_{i.}$ for the ith row may be considered separately, and

used as χ^2 with $s - 1$ degrees of freedom, or all rows may be combined into the minimized S as given in eq. (16), and used as χ^2 with $r(s - 1)$ degrees of freedom.

Solution of the two dimensional Case II. In addition to eqs. (11) we now have also

$$\sum_i \delta m_{ij} = 0 \qquad j = 1, 2, \cdots, s - 1 \tag{17}$$

which comes by differentiating eqs. (5). By addition of eqs. (10), (11), and (17), after multiplying eq. (11i) by $-\lambda_{i.}$ and eq. (17j) by $-\lambda_{.j}$, we obtain

$$\Sigma\{(m_{ij} - n_{ij})/n_{ij} - \lambda_{i.} - \lambda_{.j}\}\delta m_{ij} = 0. \tag{18}$$

Equating each brace to zero, as before, we find that

$$m_{ij} = n_{ij}(1 + \lambda_{i.} + \lambda_{.j}) \tag{19}$$

wherein $\lambda_{.s}$ is to be counted 0. The adjustment is now no longer proportionate by rows, but involves every cell.

To evaluate the Lagrange multipliers in eq. (19) we may sum the two members downward and across in Fig. 1 and obtain the $r + s - 1$ normal equations

$$\begin{aligned} n_{i.}\lambda_{i.} + \sum_j n_{ij}\lambda_{.j} &= m_{i.} - n_{i.}, \qquad i = 1, 2, \cdots, r \\ \sum_i n_{ij}\lambda_{i.} + n_{.j}\lambda_{.j} &= m_{.j} - n_{.j}, \qquad j = 1, 2, \cdots, s - 1. \end{aligned} \tag{20}$$

These can be reduced for numerical computation. The top row solved for $\lambda_{i.}$ gives

$$\lambda_{i.} = (1/n_{i.})\{m_{i.} - \sum_j n_{ij}\lambda_{.j}\} - 1 \tag{21}$$

whereupon by substitution into the bottom row of eqs. (20) we arrive at the $s - 1$ normal equations

$$\begin{array}{cccccc} \lambda_{.1} & \lambda_{.2} & \cdots & \lambda_{.s-1} & = 1 \\ \hline n_{.1} - \sum_i \frac{n_{i1}n_{i1}}{n_{i.}} & -\sum_i \frac{n_{i1}n_{i2}}{n_{i.}} & \cdots & -\sum_i \frac{n_{i1}n_{i,s-1}}{n_{i.}} & = m_{.1} - \sum_i \frac{n_{i1}m_{i.}}{n_{i.}} \\ & n_{.2} - \sum_i \frac{n_{i2}n_{i2}}{n_{i.}} & \cdots & -\sum_i \frac{n_{i2}n_{i,s-1}}{n_{i.}} & = m_{.2} - \sum_i \frac{n_{i2}m_{i.}}{n_{i.}} \\ & & & \vdots & \vdots \\ & & & n_{.s-1} - \sum_i \frac{n_{i,s-1}n_{i,s-1}}{n_{i.}} & = m_{.s-1} - \sum_i \frac{n_{i,s-1}m_{i.}}{n_{i.}} \\ & & & & 0. \end{array} \tag{22}$$

Because of symmetry in the coefficients, those below the diagonal are not shown, indeed, in a systematic computation, they are not used. The 0 in the bottom

row is appended for the computation of the minimized S, if desired. The number of Lagrange multipliers to be solved for directly is $s - 1$, and the remaining ones come by substitution into eq. (21), $\lambda_{.s}$ being counted 0.

A simple procedure for calculating the coefficients in the normal equations (22) is to set up a preparatory table by dividing each n_{ij} in the ith row by $\sqrt{n_{i.}}$; also to write down $m_{i.}/\sqrt{n_{i.}}$ for that row, for use on the right-hand side of the normal equations (compare Tables I and II). In machine calculation the constant divisor $\sqrt{n_{i.}}$ would be left on the keyboard until the entire ith row is divided; or, if reciprocal multiplication is preferred, the multiplier $1/\sqrt{n_{i.}}$ would be left on. From this preparatory table, the cumulation of squares and cross-products in the vertical gives the required summations for the coefficients. The sum check would be applied in the usual manner.

3. A numerical example of the two dimensional Case II. The fact is that in practice one need not bother about forming and solving the normal equations because they will be displaced by a simplifying iterative procedure, to be explained in a later section. For illustration, however, we may do an example both ways, first using the normal equations and the adjustment (19), later on accomplishing the same results by the quicker method.

We may start with the unitalicized numbers in the 4×6 array of Table I, assuming these to be the sampling frequencies n_{ij} to be adjusted. Actually, they were obtained by deflating 1/20th (for a supposed 5 per cent sample) the New England age $\times$ state table on p. 1108 of vol. 2 of the *Fifteenth Census of the U. S.*, 1930, then varying the deflated values by chance with Tippett's numbers to get our sampling frequencies n_{ij}. The italicized entries in Table I represent the final (adjusted) m_{ij}, and it is these that we now set out to get. We start off with the sampling frequencies n_{ij} and the known marginal totals $m_{.1}$, $m_{.2}$, etc., where $m_{i.} = N_{i.}n/N$, $m_{.j} = N_{.j}n/N$, as in eqs. (6) and (7). The Lagrange multipliers shown along the left-hand and top borders arise in the calculations now to be undertaken.

Table II is the preparatory table, advised at the close of the last section. It is derived from Table I by dividing the ith row of sample frequencies by $\sqrt{n_{i.}}$. For example, the entry 8.64 in the cell $i = 3, j = 2$ comes by dividing 419 by $\sqrt{2352}$, 419 being the entry in the cell of the same indices in Table I, and 2352 being the sum of the third row. The sums at the bottom and right-hand side are for checking the formation of the normal equations. The cumulations of squares and cross-products along the vertical give the summations required for the normal eqs. (22), which now appear numerically as eqs. (23).

	No.	$\lambda_{.1}$	$\lambda_{.2}$	$\lambda_{.3}$	=	1
	1	7413	−3549	−2354	=	3197 $\times 10^{-5}$
(23)	2		4441	−544	=	2356
	3			3129	=	−3222
	4					0

Performing the solution by any favorite procedure one will obtain

$$\lambda_{.1} = .01182 \qquad \lambda_{.2} = .01490 \qquad \lambda_{.3} = .00119 \tag{24}$$

TABLE I

A table of artificial sample frequencies, an artificial 5 percent sample of native white persons of native white parentage attending school, by age by state, New England, 1930. The adjusted frequency m_{ij} in each cell is shown italicized just below the corresponding sample frequency n_{ij}

Age			7 to 13	14 & 15	16 & 17	18 to 20	
		$j =$	1	2	3	4	$n_{i.}$
		$\lambda_{.j} =$	.0118	.0149	.0012	0	$m_{i.}$
State	i	$\lambda_{i.}$					
Maine	1	−.0146	3623	781	557	313	5274
			3613	*781*	*550*	*308*	*5252*
New Hampshire	2	−.0003	1570	395	251	155	2371
			1588	*401*	*251*	*155*	*2395*
Vermont	3	.0234	1553	419	264	116	2352
			1608	*435*	*270*	*119*	*2432*
Massachusetts	4	−.0162	10538	2455	1706	1160	15859
			10492	*2452*	*1680*	*1141*	*15766*
Rhode Island	5	−.0230	1681	353	171	154	2359
			1662	*350*	*167*	*150*	*2330*
Connecticut	6	−.0034	3882	857	544	339	5622
			3915	*867*	*543*	*338*	*5662*
		$n_{.j}$	22847	5260	3493	2237	33837
		$m_{.j}$	*22877*	*5285*	*3462*	*2213*	*33837*

The adjusted m_{ij} (italicized) are rounded off, hence when summed may occasionally disagree a unit or so with the expected marginal totals (also italicized), the latter arise by deflation from the universe rather than by direct addition of the m_{ij}.

whereupon by substitution into eq. (21) comes

$$\begin{aligned} \lambda_{1.} &= -.0146 \qquad & \lambda_{4.} &= -.0162 \\ \lambda_{2.} &= -.0003 \qquad & \lambda_{5.} &= -.0230 \\ \lambda_{3.} &= +.0234 \qquad & \lambda_{6.} &= -.0034. \end{aligned} \tag{25}$$

The next step is to compute the m_{ij} by eq. (19). Table I is now bordered with the Lagrange multipliers for a convenient arrangement of the factors required, and the calculation is completed. It will be noted that, for example

$$m_{32} = 419(1 + .0234 + .0149) = 435. \tag{26}$$

The m_{ij} thus calculated are shown italicized in Table I. The marginal totals, found by adding the m_{ij} just calculated, do not agree exactly everywhere with the expected totals, because of rounding off to integers: the errors of closure, however, are slight, and it is a simple matter to raise or lower some of the larger cells by a unit or two to force exact satisfaction of the conditions, if this is desired.

4. The three dimensional problem. Here the N cards of the universe are sorted and counted for one and perhaps a second and third characteristic, and possibly crossed by pairs in various combinations (Cases I–VII). The sample of n, however, is crossed by all three characteristics, which is to say that the

TABLE II

This comes by dividing each sample frequency in Table I by the corresponding $\sqrt{n_{i.}}$. (This operation would ordinarily be done a row at a time)

	$j=$ 1	2	3	4	$m_{i.}/\sqrt{n_{i.}}$	Sum
$i = 1$	49.89	10.75	7.67	4.31	72.32	144.94
2	32.24	8.11	5.15	3.18	49.19	97.87
3	32.02	8.64	5.44	2.39	50.15	98.64
4	83.68	19.49	13.55	9.21	125.19	251.12
5	34.61	7.27	3.52	3.17	47.97	96.54
6	51.77	11.43	7.26	4.52	75.51	150.49
Sum	284.21	65.69	42.59	26.78	420.33	839.60

cell frequencies n_{ijk} are all known (refer to Fig. 2). As before, the adjusted frequencies are required.

Case I: One set of slice totals known. Assume the slice totals $N_{1..}$, $N_{2..}$, $\cdots$, $N_{r..}$ to be known; the conditions are then

$$\sum_{jk} m_{ijk} = m_{i..} = N_{i..}n/N \qquad i = 1, 2, \cdots r \tag{27}$$

being r in number. The summation to be minimized is

$$S = \Sigma(m_{ijk} - n_{ijk})^2/n_{ijk} \tag{28}$$

being similar to that in eq. (3), except that now there are three indices to be summed over instead of two. Following a procedure similar to that used before, we differentiate eqs. (27) and (28) and introduce the r Lagrange multipliers λ_i.

with eq. (27). The steps are identical with those of the two dimensional Case I, and the result is at once

$$m_{ijk} = n_{ijk}(1 + \lambda_{i..}) = n_{ijk}(m_{i..}/n_{i..}). \tag{29}$$

This adjustment, like that shown by eq. (15), is a simple proportionate one, but this time by slices rather than by columns. All cell frequencies having the same i index are raised or lowered in the same proportion.

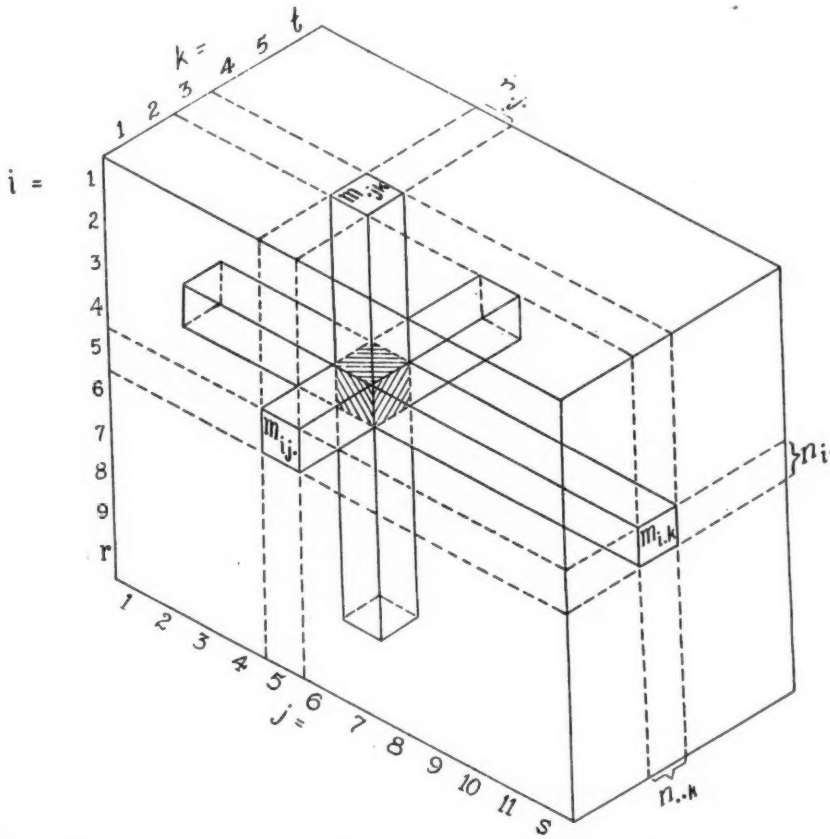


FIG. 2. SHOWING THE SYSTEM OF NOTATION FOR THE CELL FREQUENCIES AND MARGINAL TOTALS IN THE THREE DIMENSIONAL SAMPLE

Case II: Two sets of slice totals known. Here, in addition to the slice totals of Case I we know also

$$N_{.1.}, N_{.2.}, \cdots, N_{.s.}$$

whence arise the $s - 1$ additional conditions

$$\sum_{ik} m_{ijk} = m_{.j.} = N_{.j.}n/N, \qquad j = 1, 2, \cdots, s - 1. \tag{30}$$

Using the Lagrange multiplier $\lambda_{.j.}$ here, and $\lambda_{i..}$ with eq. (27) as before, we find that

$$m_{ijk} = n_{ijk}(1 + \lambda_{i..} + \lambda_{.j.}) \tag{31}$$

in which $\lambda_{.s.}$ is to be counted zero. This adjustment is proportionate by tubes, the ratio m_{ijk}/n_{ijk} being constant along the ijth tube and in fact equal to $m_{ij.}/n_{ij.}$, independent of k. Unfortunately we do not here know the face totals $m_{ij.}$ and are unable to make use of the proportionality as we shall in Case IV.

To solve for the $r + s - 1$ Lagrange multipliers we sum the members of eq. (31) over j and then over i and arrive at the normal equations

$$\begin{aligned} n_{i..}\lambda_{i..} + \sum_j n_{ij.}\lambda_{.j.} &= m_{i..} - n_{i..}, \qquad i = 1, 2, \cdots, r, \\ \sum_i n_{ij.}\lambda_{i..} + n_{.j.}\lambda_{.j.} &= m_{.j.} - n_{.j.}, \qquad j = 1, 2, \cdots, s-1. \end{aligned} \tag{32}$$

These can be reduced to $s - 1$ equations in precisely the same way that eqs. (20) were reduced, but because of the iterative process to come further on, we shall not pursue the reduction here.

Case III: All three sets of slice totals known. All slice totals

$$\begin{gathered} N_{.1.}, N_{.2.}, \cdots, N_{.s.} \\ N_{1..}, N_{2..}, \cdots, N_{r..} \\ N_{..1}, N_{..2}, \cdots, N_{..t} \end{gathered}$$

now being known, in addition to conditions (27) and (30) we require here

$$\sum_{ij} m_{ijk} = m_{..k} = N_{..k}n/N, \qquad k = 1, 2, \cdots, t-1 \tag{33}$$

which makes a total of $r + (s - 1) + (t - 1)$ or $r + s + t - 2$ conditions. The same kind of manipulation as used heretofore gives

$$m_{ijk} = n_{ijk}(1 + \lambda_{i..} + \lambda_{.j.} + \lambda_{..k}) \tag{34}$$

with $\lambda_{.s.}$ and $\lambda_{..t}$ to be counted zero. The adjustment is no longer proportionate by slices or tubes, but involves every cell. In practice, once the normal equations are solved and the Lagrange multipliers worked out, one proceeds very much as in the two dimensional Case II: for each of the t slices, corresponding to the t values of k, there will be a two dimensional adjustment, the 1 in eq. (19) being replaced now by $1 + \lambda_{..k}$.

The normal equations for the Lagrange multipliers can be found by performing double summations on eq. (34). The result is

$$\begin{aligned} n_{i..}\lambda_{i..} + \sum_j n_{ij.}\lambda_{.j.} + \sum_k n_{i.k}\lambda_{..k} &= m_{i..} - n_{i..}, \qquad i = 1, 2, \cdots, r, \\ \sum_i n_{ij.}\lambda_{i..} + n_{.j.}\lambda_{.j.} + \sum_k n_{.jk}\lambda_{..k} &= m_{.j.} - n_{.j.}, \qquad j = 1, 2, \cdots, s-1, \\ \sum_i n_{i.k}\lambda_{i..} + \sum_j n_{.jk}\lambda_{.j.} + n_{..k}\lambda_{..k} &= m_{..k} - n_{..k}, \qquad k = 1, 2, \cdots, t-1. \end{aligned} \tag{35}$$

If these calculations were to be carried out, one would simplify the computation by solving the top row for $\lambda_{i..}$, getting

$$\lambda_{i..} = (1/n_{i..})\{m_{i..} - \sum_j n_{ij.}\lambda_{.j.} - \sum_k n_{i.k}\lambda_{..k}\} - 1 \tag{36}$$

and then substituting this into the middle and last rows of eqs. (35) to get a reduced set of $s + t - 2$ normal equations for the Lagrange multipliers $\lambda_{.j.}$ and $\lambda_{..k}$, the numerical values of which when set back into eq. (36) give the $\lambda_{i..}$. In all the summations of eqs. (35) and (36), $\lambda_{.s.}$ and $\lambda_{..t}$ would be counted zero. But here again, the iterative process to be explained later will displace the use of normal equations, so actually we are not interested in reducing them.

Case IV: One set of face totals known. It may be that the rs face totals

$$N_{11.}, N_{12.}, \cdots, N_{ij.}, \cdots, N_{rs.}$$

are known from crossing the i and j characters in the universe. The conditions are then

$$\sum_k m_{ijk} = m_{ij.} = N_{ij.}n/N \qquad \begin{matrix} i = 1, 2, \cdots, r, \\ j = 1, 2, \cdots, s. \end{matrix} \tag{37}$$

The adjustment here turns out to be

$$m_{ijk} = n_{ijk}(1 + \lambda_{ij.}); \tag{38}$$

but by summing both sides over the index k to evaluate $\lambda_{ij.}$ it is seen that

$$m_{ij.} = n_{ij.}(1 + \lambda_{ij.}), \tag{39}$$

whence

$$m_{ijk} = n_{ijk}(m_{ij.}/n_{ij.}). \tag{40}$$

This adjustment is thus proportionate by tubes, like that in eq. (31), though here the factor $m_{ij.}/n_{ij.}$ is known and eq. (40) can be applied at once.

Case V: One set of face totals, and one set of slice totals known. Sometimes, in addition to the rs face totals of Case IV, the slice totals

$$N_{..1}, N_{..2}, \cdots, N_{..t}$$

will also be known, in which circumstances the conditions (37) are to be accompanied by

$$\sum_{ij} m_{ijk} = m_{..k} = N_{..k}n/N, \qquad k = 1, 2, \cdots, t - 1. \tag{41}$$

The same procedure as previously applied yields now

$$m_{ijk} = n_{ijk}(1 + \lambda_{ij.} + \lambda_{..k}) \tag{42}$$

with $\lambda_{..t}$ to be counted zero. Summations performed over k, and then over i and j together, give the normal equations

$$
(43)\qquad
\begin{aligned}
n_{ij.}\lambda_{ij.} + \sum_k n_{ijk}\lambda_{..k} &= m_{ij.} - n_{ij.}, \\
\sum_{ij} n_{ijk}\lambda_{ij.} + n_{..k}\lambda_{..k} &= m_{..k} - n_{..k}.
\end{aligned}
$$

The number of equations is $rs + t - 1$, since $\lambda_{..t}$ does not exist. As before, a simplification can be effected by solving the top row for $\lambda_{ij.}$ and making a substitution into the lower one, but because of the great advantage of the iterative process to be seen further on, we shall not carry out the reduction.

Before going on it might be noted that although this case is three dimensional, it reduces to the two dimensional Case II if one considers that $ij.$ is one index running through the values $11, 12, \cdots, 21, 22, \cdots, rs$, and that $..k$ is a second index running through the values $1, 2, \cdots, t$. This can be seen by the similarity between eqs. (43) and (20).

Case VI: Two sets of face totals known. If in addition to the face totals of Case IV, the face totals

$$N_{.11}, N_{.12}, \cdots, N_{.st}$$

are also known from further crossing the j and k characters in the universe, we shall require

$$
(44)\qquad \sum_i m_{ijk} = m_{.jk} = N_{.jk}n/N, \qquad
\begin{aligned} j &= 1, 2, \cdots, s, \\ k &= 1, 2, \cdots, t-1 \end{aligned}
$$

in addition to the conditions (37). In place of eq. (40) of Case IV we now find that

$$(45)\qquad m_{ijk} = n_{ijk}(1 + \lambda_{ij.} + \lambda_{.jk})$$

in which $\lambda_{.jt}$ is to be counted zero for all j. No simple relation such as eq. (40) is possible here, because the adjustment is not proportionate by tubes; the Lagrange multipliers must be evaluated. This can be accomplished by summing the members of eq. (45) over k and i in turn, resulting in the normal equations

$$
(46)\qquad
\begin{aligned}
n_{ij.}\lambda_{ij.} + \sum_k n_{ijk}\lambda_{.jk} &= m_{ij.} - n_{ij.}, \\
\sum_i n_{ijk}\lambda_{ij.} + n_{.jk}\lambda_{.jk} &= m_{.jk} - n_{.jk}.
\end{aligned}
$$

Since $\lambda_{.jt}$ does not exist for any values of j, the number of equations is $rs + s(t - 1) = s(r + t - 1)$. They break up at once into s sets each of $r + t - 1$ equations, one set for every j value. In fact, the problem can be considered as s sets of the two dimensional Case II. Any one value of j gives a slice, which can be looked upon as fulfilling the specifications of the two dimensional Case II. Each set of normal equations can be reduced in the same manner that eqs. (20) were reduced.

Case VII: All three sets of face totals known. All totals now being known, we require

$$\sum_k m_{ijk} = m_{ij.} = N_{ij.}n/N, \qquad \begin{aligned} i &= 1, 2, \cdots, r, \\ j &= 1, 2, \cdots, s, \end{aligned} \tag{37}$$

$$\sum_i m_{ijk} = m_{.jk} = N_{.jk}n/N, \qquad \begin{aligned} j &= 1, 2, \cdots, s, \\ k &= 1, 2, \cdots, t-1, \end{aligned} \tag{44}$$

$$\sum_j m_{ijk} = m_{i.k} = N_{i.k}n/N, \qquad \begin{aligned} i &= 1, 2, \cdots, r-1, \\ k &= 1, 2, \cdots, t-1. \end{aligned} \tag{47}$$

The adjusting relation is

$$m_{ijk} = n_{ijk}(1 + \lambda_{ij.} + \lambda_{.jk} + \lambda_{i.k}) \tag{48}$$

in which $\lambda_{.jt}$ is to be counted zero for any j, $\lambda_{r.k}$ for any k, and $\lambda_{i.t}$ for any i. The normal equations for the Lagrange multipliers are

$$\begin{aligned} n_{ij.}\lambda_{ij.} + \sum_k n_{ijk}\lambda_{.jk} + \sum_k n_{ijk}\lambda_{i.k} &= m_{ij.} - n_{ij.} \\ \sum_i n_{ijk}\lambda_{ij.} + n_{.jk}\lambda_{.jk} + \sum_i n_{ijk}\lambda_{i.k} &= m_{.jk} - n_{.jk} \\ \sum_j n_{ijk}\lambda_{ij.} + \sum_j n_{ijk}\lambda_{.jk} + n_{i.k}\lambda_{i.k} &= m_{i.k} - n_{i.k} \end{aligned} \tag{49}$$

being $rs + rt + st - r - s - t + 1$ in number. They can be reduced in the same way that previous normal equations have been reduced; but here again, the iterative process will render the use of normal equations unnecessary, except for theoretical purposes, e.g. justification of the iterative process.

5. A simplified procedure—iterative proportions. It is well known in least squares that the number of Lagrange multipliers in any problem is equal to the number of conditions imposed on the adjustment. Here the conditions have appeared in sets, depending on which marginal totals are involved. By a comparison of eqs. (15) and (29) on the one hand, with eqs. (19), (31), (34), (42), (45), and (48) on the other, we see that wherever there was only one set of marginal totals involved we came out with a proportionate adjustment, but that in all other cases it was not so; the Lagrange multipliers involved were unfortunately related to one another through normal equations. We now make the observation, however, that as a first approximation the adjustments may all be considered proportionate, and we shall be able to write down an expression for the error in this approximation, and shall be able to eliminate it by a succession of proportionate adjustments.

Take the two dimensional Case II for an example. In eq. (21) one may recognize $(1/n_{i.}) \sum_j n_{ij}\lambda_{.j}$ as a weighted average of $\lambda_{.j}$ for the ith row. There will be a weighted average of $\lambda_{.j}$ for the first row, another for the second, etc., one for each value of i; consequently one may appropriately speak of the ith

average of $\lambda_{.j}$, writing it i-av. $\lambda_{.j}$. Substituting from eq. (21) into (19) one then sees the adjustment (19) appear as

$$m_{ij} = n_{ij}(m_{i.}/n_{i.} + \lambda_{.j} - i\text{-av. } \lambda_{.j}). \tag{50}$$

If, on the other hand, $\lambda_{.j}$ had been eliminated from eqs. (20), instead of $\lambda_{i.}$, the result would have been

$$m_{ij} = n_{ij}(m_{.j}/n_{.j} + \lambda_{i.} - j\text{-av. } \lambda_{i.}). \tag{51}$$

From either eq. (50) or (51) it is clear why the adjustment (19) is not proportionate by rows or columns, and why Case II does not break up into r or s sets of Case I: the reason is that $\lambda_{.j}$ in any cell is not necessarily equal to the average $\lambda_{.j}$ for that row, nor is $\lambda_{i.}$ in any cell necessarily equal to the average $\lambda_{i.}$ for that column. If nevertheless one were to make the simple proportionate adjustment

$$m'_{ij} = n_{ij}(m_{i.}/n_{i.}) \tag{52}$$

along the horizontal in the ith row, the horizontal conditions (4) will be enforced but not the vertical ones (5); i.e., it will be found that $m'_{i.} = m_{i.}$, but that usually not all $m'_{.j} = m_{.j}$. This is because eq. (52) effects only a partial adjustment, each m'_{ij} being in error through the disparity between the $\lambda_{.j}$ proper to the jth column, and the average of all the $\lambda_{.j}$ for the ith row, as seen in eq. (50). This error can then be diminished by turning the process around and subjecting these m'_{ij} to a proportionate adjustment in the vertical according to the equation

$$m''_{ij} = m'_{ij}(m_{.j}/m'_{.j}) \tag{53}$$

which may be considered an application of eq. (51) wherein the disparity between any $\lambda_{i.}$ and the average $\lambda_{i.}$ for the jth column has been neglected. It is the vertical conditions that will now be found satisfied, but perhaps not all of the horizontal ones, because some of the row totals may have been disturbed. The cycle initiated by eq. (52) is therefore repeated, and the process is continued until the table reproduces itself and becomes rigid with the satisfaction of all the conditions, both horizontal and vertical. The final results coincide with the least squares solution, which is thus accomplished without the use of normal equations.

Usually two cycles suffice. In practice the work proceeds rapidly, requiring only about one-seventh as much time as setting up the normal equations and solving them. The tables III–V show the various stages of the work when the method of iterative proportions is applied to the sample frequencies of Table I. It will be noticed that the results of the third approximation (Table V) are final, since if the process were continued, the table would only reproduce itself.

The same process can be extended to three or more dimensions with an even greater relative saving in time. To see how the method of iterative proportions

applies in one of the three dimensional cases, we may go back to Case III. By the substitution afforded through eq. (36) the adjusting eq. (34) may be put into the form

TABLE III

The method of iterative proportions applied to the data of Table I. First stage: A proportionate adjustment by rows by eq. (52). Note that $m'_{i.} = m_{i.}$, *but that* $m'_{.j} \neq m_{.j}$

	$j = 1$	2	3	4	m'	m_i
$i = 1$	3608	778	555	312	5253	5252
2	1586	399	254	157	2396	2395
3	1606	433	273	120	2432	2432
4	10476	2441	1696	1153	15766	15766
5	1660	349	169	152	2330	2330
6	3910	863	548	341	5662	5662
$m'_{.j}$	22846	5263	3495	2235	33839	
$m_{.j}$	22877	5285	3462	2213		33837

TABLE IV

A continuation of the process initiated in Table III. The figures in Table III are now adjusted proportionately by columns according to eq. (53). The vertical totals $m''_{.j}$ *and* $m_{.j}$ *now are equal, but the agreement of the horizontal totals accomplished in Table III has been slightly disturbed*

	$j = 1$	2	3	4	$m''_{i.}$	$m_{i.}$
$i = 1$	3613	781	550	309	5253	5252
2	1588	401	252	155	2396	2395
3	1608	435	270	119	2432	2432
4	10490	2451	1680	1142	15763	15766
5	1662	350	167	151	2330	2330
6	3915	867	543	338	5663	5662
$m''_{.j}$	22876	5285	3462	2214	33837	
$m_{.j}$	22877	5285	3462	2213		33837

(54) $$m_{ijk} = n_{ijk}(m_{i..}/n_{i..} + \lambda_{.j.} + \lambda_{..k} - i\text{-av. } \lambda_{.j.} - i\text{-av. } \lambda_{..k}).$$

Equally well it could have been written

(55) $$m_{ijk} = n_{ijk}(m_{.j.}/n_{.j.} + \lambda_{i..} + \lambda_{..k} - j\text{-av. } \lambda_{i..} - j\text{-av. } \lambda_{..k}),$$

or

$$m_{ijk} = n_{ijk}(m_{..k}/n_{..k} + \lambda_{i..} + \lambda_{.j.} - k\text{-av. } \lambda_{i..} - k\text{-av. } \lambda_{.j.}). \tag{56}$$

Any of these three equations shows why the adjustment (34) is not proportional by slices, and why this case does not break up into r or s or t sets of the three dimensional Case I. As a first approximation it does, as is now clear from these three equations, and by making successive proportionate adjustments we may thus arrive at the least squares values. To go about the work we could first calculate the values of

$$m'_{ijk} = n_{ijk}(m_{i..}/n_{i..}) \tag{57}$$

then

$$m''_{ijk} = m'_{ijk}(m_{.j.}/m'_{.j.}) \tag{58}$$

TABLE V

The cycle is commenced again. The figures of Table IV are subjected to a proportionate adjustment by rows, according to eq. (52). And since these results turn out to be almost a reproduction of Table IV but with both horizontal and vertical conditions satisfied, they are considered final. The agreement with the m_{ij} in Table I should be noted

	$j = 1$	2	3	4	$m'_{i.}$	$m_{i.}$
$i = 1$	3612	781	550	309	5252	5252
2	1587	401	252	155	2395	2395
3	1608	435	270	119	2432	2432
4	10492	2451	1680	1142	15765	15766
5	1662	350	167	151	2330	2330
6	3914	867	543	338	5662	5662
$m'_{.j}$	22875	5285	3462	2214	33836	
$m_{.j}$	22877	5285	3462	2213		33837

followed by

$$m'''_{ijk} = m''_{ijk}(m_{..k}/m''_{..k}). \tag{59}$$

These three successive adjustments would constitute a cycle, which would then be repeated in whole or in part until the table becomes rigid with the satisfaction of all three sets of conditions.

6. Simplification when only one cell requires adjustment. On occasions it happens in sampling work that one is especially interested in one particular cell of the universe, and would like to have a result for it in advance before the other cells are adjusted. Sometimes it even happens that the others individually are of no particular concern. In such circumstances one merely places the cell

of interest in one corner of the table by an appropriate interchange of rows and columns, and then compresses the rest of the table into the cells adjacent to it. In the two dimensional Case II one would thus work with a 2×2 table, one corner cell being the one of special interest, the other three being the result of compression. The marginal totals of the row and column belonging to the cell of interest are unaffected. For illustration we may suppose that from the sample shown in Table I we require only m_{61}. We then start with the 2×2 Table VI, which is derived from Table I by compression. Commencing with Table VI, one might first adjust by rows according to eq. (52), then by columns by eq. (53). One cycle of iterative proportions is sufficient, as is seen in Table

TABLE VI

Derived from Table I by compression, the cell $i = 6, j = 1$, requiring adjustment

	$j = 1$	$j = 2 - 4$	$n_{i.}$	$m_{i.}$
$i = 1 - 5$	18965	9250	28215	28175
$i = 6$	3882	1740	5622	5662
$n_{.j}$	22847	10990	33837	
$m_{.j}$	22877	10960		33837

TABLE VII

A proportionate adjustment of Table VI

Rows adjusted by eq. (52)

18938	9237	28175
3910	1752	5662
22848	10989	33837

Columns adjusted by eq. (53)

18962	9213	28175
3915	1747	5662
22877	10960	33837

Conclusion: $m_{61} = 3915$

VII, and the value 3915 found for m_{61} is in good agreement with its value shown in Tables I and V. The scheme of compression provides a quick method of getting out an advance adjustment for a cell of special interest, and the result so obtained will ordinarily be in good agreement with what comes later when and if all the cells are adjusted.

In the three dimensional Cases II, III, V, VI, and VII, one compresses the original table to a $2 \times 2 \times 2$ table, and then uses the method of iterative proportions. (The other cases do not require consideration, since they are proportionate adjustments wherein one is already at liberty to adjust as few or as many cells as he likes without altering the equations or the routine.) The same procedure can be extended to the adjustment of two cells, the only modification

being that in two dimensions we shall compress to a 2×3 or a 3×3 table, depending on whether the two cells do or do not lie in the same row or column. In three dimensions we compress to a $2 \times 2 \times 3$, or a $2 \times 3 \times 3$, or a $3 \times 3 \times 3$ table; the first if the two cells lie in the same i, j, or k tube, the second if they lie in the same slice but not in the same tube, the third if they are in separate slices.

7. Some remarks on the accuracy of an adjustment. A least squares adjustment of sampling results must be regarded as a systematic procedure for obtaining satisfaction of the conditions imposed, and at the same time effecting an improvement of the data in the sense of obtaining results of smaller variance than the sample itself, under ideal conditions of sampling from a stable universe. It must not be supposed that any or all of the adjusted m_{ij} in any table are necessarily "closer to the truth" than the corresponding sampling frequencies n_{ij}, even under ideal conditions. As for the standard errors of the adjusted results, they can easily be estimated for the ideal case by making use of the calculated chi-square. For predictive purposes, however (which can be regarded as the only possible use of a census by any method, sample or complete), it is far preferable, in fact necessary, to get some idea of the errors of sampling by actual trial, such as by a comparison of the sampling results with the universe, as can often be arranged by means of controls. There is another aspect to the problem of error—even a 100 per cent count, even though strictly accurate, is not by itself useful for prediction, except so far as we can assert on other grounds what secular changes are taking place.

In conclusion it is a pleasure to record our appreciation of the assistance of Miss Irma D. Friedman and Mr. Wilson H. Grabill for putting the formulas and procedure into actual operation with census data, and thereby disclosing defects in earlier drafts of the manuscript.

BUREAU OF THE CENSUS,
WASHINGTON

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

THE STANDARD ERRORS OF THE GEOMETRIC AND HARMONIC MEANS AND THEIR APPLICATION TO INDEX NUMBERS[1]

By Nilan Norris

Attempts to derive useful expressions for estimating the standard deviations of the sampling errors of the geometric and harmonic means have not yielded results comparable with those afforded by the modern theory of estimation, including fiducial inference. There are in the literature of probability theory certain theorems which can be applied to obtain these desired results in a straightforward manner. The use of forms for estimating standard errors is subject to certain conditions which are not always fulfilled, particularly in the case of time series. An understanding of these limitations should deter those who may be tempted to judge the significance of phenomena such as price changes solely on the basis of estimated standard errors of indexes.

1. Statement of formulas. The standard error of the geometric mean of a sequence of positive independent chance variables denoted by $x_i = x_1, x_2, \cdots, x_n$, is $\sigma_G = \theta_1 \dfrac{\sigma_{\log x}}{\sqrt{n}}$, where θ_1 is the population geometric mean of the variates; so that $\sigma_{\log x}$ is the standard deviation of the logarithms in the population as given by $\sigma_{\log x} = [E\{[\log x - E(\log x)]^2\}]^{\frac{1}{2}}$; and n is the number of individuals comprising the sample. The estimate of the standard error of the geometric mean is $s_G = G \dfrac{s_{\log x_i}}{\sqrt{n-1}}$, where G is the sample geometric mean, that is, the estimate of θ_1; so that $s_{\log x_i}$ is the estimate of $\sigma_{\log x}$; and $n - 1$ is the degree of freedom of the sample.

---

[1] This article summarizes two papers presented at sessions of the Institute of Mathematical Statistics at Detroit, Michigan on December 27, 1938, and at Philadelphia, Pennsylvania on December 27, 1939. The results given herein can be derived by several methods, which vary somewhat as to degree of rigor. The writer wishes to acknowledge his indebtedness to the referee for suggesting a proof based on a probability theorem stated by J. L. Doob, "The limiting distributions of certain statistics," *Annals of Math. Stat.*, Vol. 4 (1935), pp. 160–169. The standard deviation formulas obtained follow as an application of this theorem, as will be seen by reference to it. Obviously the asymptotic variance formulas of many other statistics (estimates of parameters) can be obtained in a similar manner.

The standard error of the harmonic mean of a sequence of positive independent chance variables denoted by $x_i = x_1, x_2, \cdots, x_n$, is $\sigma_H = \theta_2^2 \frac{\sigma_{1/x}}{\sqrt{n}}$, where the population harmonic mean of the variates is $\theta_2 = 1/\alpha = [E(1/x)]^{-1}$; so that the standard deviation of $1/x$ in the population is $\sigma_{1/x} = [E\{[1/x - E(1/x)]^2\}]^{\frac{1}{2}}$; and n is the number of observations comprising the sample. The estimate of the standard error of the harmonic mean is $s_H = \frac{1}{a^2} \frac{s_{1/x_i}}{\sqrt{n-1}}$, where the estimate of α is given by $a = \frac{1}{H} = \frac{1}{n}(\Sigma\, 1/x_i)$; in which s_{1/x_i} is the standard deviation of the reciprocals of the observations comprising the sample; and $n - 1$ is the degree of freedom of the sample.

2. Derivation of formulas. These forms can be obtained by application of the Laplace-Liapounoff theorem[2] as follows: Let $x_i = x_1, x_2, \cdots, x_n$ be a set of positive independent chance variables with the same distribution functions, where the expectations, $E(x_i)$ and $E(x_i^2)$ exist, and where $\sigma_x^2 = E\{[x_i - E(x_i)]^2\} > 0$. The last condition is imposed to eliminate the trivial case in which the x_i are all equal and their distribution is confined to a single point. The geometric mean of the x_i is $G = (x_1 \cdot x_2 \cdot \cdots \cdot x_n)^{1/n}$, and the harmonic mean of the x_i is $H = \left[\frac{1}{n}\sum \frac{1}{x_i}\right]^{-1}$.

It is necessary to assume that both $\sigma_{\log x}^2$ and $\sigma_{1/x}^2$ are finite, and that in the case of both $\log x$ and $1/x$ at least one moment of order higher than any two of the respective variates is also finite. The requirement that the variance and at least one moment higher than the variance be finite can be weakened in various ways, but this is a trivial consideration, since nearly all distributions of any importance have finite third moments.[3] Certain rarely occurring types of distributions, such as the Cauchy distribution, have infinite variance. In such cases, standard error formulas as ordinarily used are not valid.

Let $E(\log x) = \zeta$, and $E(1/x) = \alpha$. By the Laplace-Liapounoff theorem, except for terms of order $1/\sqrt{n}$, the limiting distributions of $\frac{\sqrt{n}(\log G - \zeta)}{\sigma_{\log x}}$ and $\frac{\sqrt{n}(H^{-1} - \alpha)}{\sigma_{1/x}}$ are normal with zero arithmetic means and unit variances. That is, if C represents a set of conditions on chance variables, and $P\{C\}$ is the probability that these conditions are satisfied, then

---

[2] A. Khintchine, Asymptotische Gesetze der Wahrscheinlichkeitsrechnung, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, J. Springer, Berlin, 1933, Vol. II, No. 4, pp. 1–8; J. L. Doob, *op. cit.*, pp. 160–169; and S. S. Wilks, *Statistical Inference*, 1936–1937, Edwards Brothers, Inc., Ann Arbor, 1937, pp. 39 *f.*

[3] For a more detailed discussion of this matter see Wilks, *op. cit.*, pp. 39 *f.*

$$\lim_{n\to\infty} P\left\{\frac{\sqrt{n}(\log G - \zeta)}{\sigma_{\log x}} < t\right\} = \lim_{n\to\infty} P\left\{\frac{\sqrt{n}(H^{-1} - \alpha)}{\sigma_{1/x}} < t\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}}\, dx.$$

In order to use these relations in obtaining the limiting distributions of the geometric and harmonic means, it is necessary to suppose that the sequence of random chance variables, V_i, converges in probability (converges stochastically) to ρ, and that the sequence of random chance variables, $\sqrt{n}(V_i - \rho)$, has a normal limiting distribution with zero arithmetic mean and variance σ^2. Also, it is necessary to assume that the real-valued function, $f(x)$, has a Taylor expansion valid in the neighborhood of ρ. If $f'(\rho) \neq 0$, only the first two terms of the series are needed. The required expansion is given by

$$f(x) = f(\rho) + (x - \rho)f'(\rho) + \frac{(x - \rho)^2}{2} f''[\rho + \beta(x - \rho)],$$

where $0 < \beta < 1$, and $f''(x)$ is continuous in the neighborhood of ρ. When these conditions are fulfilled, the limiting distribution of $\sqrt{n}[f(V_i) - f(\rho)]$ is normal with an arithmetic mean of zero and a variance of $\sigma^2[f'(\rho)]^2$.

Let $f(\log G) = e^{\log G}$, and use the expansion given by $e^{\log G} = e^{\zeta} + (\log G - \zeta)e^{\zeta} + \frac{1}{2}(\log G - \zeta)^2\, e^{\zeta + \beta(\log G - \zeta)}$. Since $\theta_1 = e^{\zeta}$, it follows that the limiting distribution of $\sqrt{n}(G - \theta_1)$ is normal with an arithmetic mean of zero and a variance of $\theta_1^2\sigma_{\log x}^2$.

Similarly, it can be shown that the limiting distribution of $\sqrt{n}(H - \theta_2)$ is normal with an arithmetic mean of zero and a variance of $\theta_2^4\sigma_{1/x}^2$, where $\theta_2 = \frac{1}{\alpha} = [E(1/x)]^{-1}$.

It is of some interest to observe that the expressions for the standard errors of the geometric and harmonic means correspond with the forms previously given for the standard errors of two efficient ratio-measures of relative variation,[4] namely,

$$\sigma_{G/A} = \frac{\theta_1^2}{\theta^2}\, \sigma_{A/G}, \quad \text{and} \quad \sigma_{H/G} = \frac{\theta_2^2}{\theta_1^2}\, \sigma_{G/H},$$

where θ_1/θ is the population geometric-arithmetic ratio, and θ_2/θ_1 is the population harmonic-geometric ratio.

3. Limitations of standard-error estimates. Application of these forms is subject to the usual conditions for drawing sound inferences on the basis of the representative method. Fiducial argument should be employed to avoid certain untenable assumptions of the outmoded method of using standard errors. Estimates of the standard deviations of sampling errors do not constitute an ultimate test of significance which can be applied with a high degree of success to all types of problems. In general, such estimates cannot be relied upon with a

[4] Nilan Norris, "Some efficient measures of relative dispersion," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 214–220.

high degree of confidence when they are used as tests of significance for index numbers, since in nearly all time series there exists an appreciable degree of serial correlation, persistence, or lack of independence among successive items of any sample.

4. Bibliographical note. Certain aspects of the sampling distribution of the geometric mean have been discussed by Burton H. Camp.[5] Attempts to derive forms for estimating the standard errors of index numbers have been made by Truman L. Kelley[6] and Irving Fisher,[7] and an empirical study of the sampling fluctuations of indexes has been made by E. C. Rhodes.[8] Although various special tests of significance for time series have been proposed,[9] at the present time no generally satisfactory procedure has appeared.

HUNTER COLLEGE,
NEW YORK, N. Y.

[5] Burton H. Camp, "Notes on the distribution of the geometric mean," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 221–226.

[6] Truman L. Kelley, "Certain Properties of Index Numbers," *Quarterly Publications of Am. Stat. Assn.*, Vol. 17, New Series 135, Sept., 1921, pp. 826–841.

[7] Irving Fisher, *The Making of Index Numbers*, Houghton Mifflin Company, New York, 1927, 3d ed., pp. 225–229, 342–345, and Appendix I, pp. 407 and 430 *f*.

[8] E. C. Rhodes, "The precision of index numbers," *Roy. Stat. Soc. Jour.*, Vol. 99 (1936), Part I, pp. 142–146, and Part II, pp. 367–369.

[9] Some of the more recent papers dealing with this matter are: G. Tintner, "On tests of significance in time series," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 139–143; "The analysis of economic time series," *Am. Stat. Assn. Jour.*, Vol. 35 (1940), pp. 93–100; L. R. Hafstad, "On the Bartels technique for time-series analysis, and its relation to the analysis of variance," *Am. Stat. Assn. Jour.*, Vol. 35 (1940), pp. 347–361; and Lila F. Knudsen, "Interdependence in a series," *Am. Stat. Assn. Jour.*, Vol. 35 (1940), pp. 507–514.

A NOTE ON THE USE OF A PEARSON TYPE III FUNCTION IN RENEWAL THEORY

BY A. W. BROWN

One of the methods suggested by A. J. Lotka[1] for the derivation of the renewal function may be briefly summarized as follows.

The method consists of dissecting the total renewal function into "generations". The original installation constitutes the zero generation, the units introduced to replace disused units of the zero generation constitute the first generation, renewal of these the second, and so on. Let $f(x)$ be the "mortality" function, the same for all generations. $f(x)$ is a function satisfying the usual conditions of a distribution function. Adopting Lotka's notation, let N be the number of units in the original collection, $B_1(t)\,dt$ the number of objects intro-

[1] A. J. Lotka, "A Contribution to the Theory of Self Renewing Aggregates, With Special Reference to Industrial Replacement," *Annals of Math. Stat.*, Vol. 10 (1939), p. 1.

duced between times t and $t + dt$ and belonging to the first generation, $B_2(t)\,dt$ a similar expression for the second generation, etc. $B_1(t)/N$, $B_2(t)/N, \ldots$ may be regarded as renewal density functions for the various generations.

Now, evidently,

$$B_1(t) = Nf(t) \tag{1}$$

$$B_2(t) = \int_0^t B_1(t - x)f(x)\,dx \tag{2}$$

and in general

$$B_{j+1}(t) = \int_0^t B_j(t - x)f(x)\,dx. \tag{3}$$

Summation of the contributions of the successive generations gives for the total renewal at the time t

$$B(t) = B_1(t) + \int_0^t B(t - x)f(x)\,dx. \tag{4}$$

In this note we propose to use a Pearson Type III function for $f(x)$ and observe what form our equations then assume. The Pearson Type III function $\frac{c^k}{\Gamma(k)}\, x^{k-1}e^{-cx}$, $(c > 0,\ k > 0)$, appears to be a reasonable one to use in many practical situations. The two parameters c and k give it a considerable amount of flexibility. The fact that this function has an unlimited range in one direction is relatively unimportant from a practical point of view, as is well known from the experience of fitting curves of this type to skewed data with limited range. Of course the question of whether a Type III curve is appropriate can be answered more objectively by using the usual Pearson curve-fitting criteria, β_1, β_2 and k. We have, then, substituting in (1)

$$B_1(t) = N\,\frac{c^k}{\Gamma(k)}\, t^{k-1}e^{-ct} \tag{5}$$

and from (2)

$$B_2(t) = \int_0^t N\,\frac{c^k}{\Gamma(k)}\,(t - x)^{k-1}e^{-c(t-x)}\,\frac{c^k}{\Gamma(k)}\,x^{k-1}e^{-cx}\,dx \tag{6}$$

$$= \frac{Nc^{2k}}{\Gamma(k)\Gamma(k)}\,e^{-ct}\int_0^t (t - x)^{k-1}x^{k-1}\,dx. \tag{7}$$

If, now, we set $x = ty$, the integral in (7) reduces to

$$\int_0^t (t - x)^{k-1}x^{k-1}\,dx = t^{2k-1}\,\frac{\Gamma(k)\Gamma(k)}{\Gamma(2k)}.$$

Hence,

$$B_2(t) = N \frac{c^{2k}}{\Gamma(2k)} t^{2k-1} e^{-ct} \tag{8}$$

and in general

$$B_j(t) = N \frac{c^{jk}}{\Gamma(jk)} t^{jk-1} e^{-ct}. \tag{9}$$

Summing the contributions of the several generations, we have for the total renewal function

$$B(t) = Nce^{-ct} \left\{ \frac{(ct)^{k-1}}{\Gamma(k)} + \frac{(ct)^{2k-1}}{\Gamma(2k)} + \cdots \right\}. \tag{10}$$

If k is a positive integer ≥ 3, (10) can be easily summed to a form which shows immediately its damped periodic nature. Even if k is positive but not an integer, it can be shown by continuity considerations that the function $B(t)$ defined by (10) has periodic properties.

Assuming k to be a positive integer, then, and setting $z = ct$, we may write the expression in brackets in (10) as

$$\frac{z^{k-1}}{(k-1)!} + \frac{z^{2k-1}}{(2k-1)!} + \cdots = f(z). \tag{11}$$

Then

$$\frac{d^k f(z)}{dz^k} = f(z)$$

and upon making the trial substitution, $f(z) = Ae^{mz}$, we get

$$Am^k e^{mz} = Ae^{mz}.$$

Hence,

$$m^k = 1.$$

Taking unity in its complex form

$$1 = \cos 2n\pi + i \sin 2n\pi$$

we have that

$$m_n = \sqrt[k]{1} = \cos \frac{2n\pi}{k} + i \sin \frac{2n\pi}{k} \tag{12}$$

where $n = 0, 1, 2, \cdots, k - 1$. Then

$$f(z) = \sum_{n=0}^{k-1} A_n e^{m_n z}$$

and

$$f^j(z) = \sum_{n=0}^{k-1} A_n m_n^j e^{m_n z}.$$

Now setting $z = 0$, we get

$$f(0) = A_0 + A_1 + \cdots + A_{k-1} = 0$$
$$f'(0) = A_0 m_0 + A_1 m_1 + \cdots + A_{k-1} m_{k-1} = 0$$
$$\vdots$$
$$f^{k-1}(0) = A_0 m_0^{k-1} + A_1 m_1^{k-1} + \cdots + A_{k-1} m_{k-1}^{k-1} = 1$$

k equations to determine the k constants. We know that A_n is equal to the ratio of two determinants formed from the coefficients of the above equations. This ratio reduces to

$$A_n = \frac{(-1)^{k+n+1}}{(m_{k-1} - m_n)(m_{k-2} - m_n) \cdots (m_n - m_0)}. \tag{13}$$

We have, then, an expression for the k constants in terms of the k roots of unity. Therefore, for any particular value of k we can obtain the sum of our series from the relation

$$f(z) = \sum_{n=0}^{k-1} A_n e^{m_n z}.$$

Hence, under the assumption that k is a positive integer, we have

$$B(t) = Nce^{-ct} \sum_{n=0}^{k-1} A_n e^{m_n ct}. \tag{14}$$

The forms of $B(t)$ for $k = 1, 2, 3, 4$ are respectively

$$B(t) = Nc$$
$$B(t) = \tfrac{1}{2}Nc(1 - e^{-2ct})$$
$$B(t) = Nce^{-ct}\left[\tfrac{1}{3}e^{ct} - e^{-\frac{1}{2}ct}\left(\tfrac{1}{3}\cos \tfrac{1}{2}\sqrt{3}ct + \frac{1}{\sqrt{3}}\sin \tfrac{1}{2}\sqrt{3}ct\right)\right]$$
$$B(t) = Nce^{-ct}[\tfrac{1}{4}(e^{ct} - e^{-ct}) - \tfrac{1}{2}\sin ct].$$

Although the above procedure is valuable particularly because it brings to light something of the nature of our renewal function, the forms derived above can be used actually to obtain values of $B(t)$ for various values of t. However, for extensive numerical work a better method is at hand, which does not even depend on the assumption of an integral value for k.

Let us return once again to equation (10) which may be written in the following form

$$B(t) = Nc\left\{\frac{e^{-ct}(ct)^{k-1}}{\Gamma(k)} + \frac{e^{-ct}(ct)^{2k-1}}{\Gamma(2k)} + \cdots\right\}. \tag{15}$$

If k and c are determined by the method of moments, (using two moments), k will not, in general, be a positive integer. However, by using the *Tables of the Incomplete Gamma Function* edited by Karl Pearson, one can compute values of $B(t)$ without much difficulty. In these tables the function $I(u, p)$ is tabulated for various values of u and p, where $I(u, p)$ is defined by

$$I(u, p) = \frac{\int_0^{u\sqrt{p+1}} e^{-v} v^p \, dv}{\Gamma(p + 1)}. \tag{16}$$

If we let $\xi = u_1\sqrt{p + 1} = u_0\sqrt{p}$ then upon integrating by parts we find

$$\frac{e^{-\xi} \xi^p}{\Gamma(p + 1)} = I(u_0, p - 1) - I(u_1, p). \tag{17}$$

The left hand member of this equation is of the same form as each of the terms of the series in brackets in (15). Hence, the value of the renewal function for a particular time, t, is directly obtainable by summation of the right hand member of (17) for successive significant values of the argument p.

By way of illustration a numerical example will be considered. The data are taken from E. B. Kurtz' book entitled *Life Expectancy of Physical Property.* In this book the author makes a study of retirement rates of fifty-two different types of physical property, and finds that their replacement curves fall into seven distinct groups. We consider here Group VII which happens to be the largest group, embracing seventeen different types of industrial equipment out of the fifty-two examined. Using Kurtz' replacement data [2] we obtain for the value of the first and second moments

$$\mu_1 = 10.002$$

$$\mu_2 = 121.71$$

and from these by the method of moments, we find

$$k = 4.62$$

$$c = .462.$$

We then proceed to calculate values of $B(t)/N$ by means of Pearson's Tables,[3] obtaining the results shown in the following table.

---

[2] E. B. Kurtz, *Life Expectancy of Physical Property,* Ronald Press, 1930, Table 22, page 86.

[3] With regard to the method of interpolation employed in the calculations, it should be mentioned that it was found advisable to use the Mid-panel Central Difference Formula (*xxiii*) on page *xii* of the introduction to Pearson's Tables; and that it is quite sufficient for our purposes to calculate only first order terms.

t	$B(t)/N$	t	$B(t)/N$
0	.0000	10	.1049
1	.0016	11	.1043
2	.0103	12	.1028
3	.0279	13	.1006
4	.0486	14	.0990
5	.0714	15	.0994
6	.0867	16	.1009
7	.0980	17	.1013
8	.1039	18	.0992
9	.1066	19	.0999
		20	.0993

In conclusion the author wishes to thank Professor S. S. Wilks for various suggestions he has made in connection with this note.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

ESTIMATES OF PARAMETERS BY MEANS OF LEAST SQUARES

BY EVAN JOHNSON, JR.

As a criterion for comparing estimates of a parameter of a universe, of known type of distribution, the use of the principle of least squares is suggested. A criterion may be stated in rather general terms. Its application to any given problem presumes a knowledge of the distribution functions of the estimates considered. In the present paper a criterion is set up and application of it is made in the estimation of the mean and of the square of standard deviation of a normal universe.

We shall use the symbol θ to represent a parameter to be estimated. It is to be remembered that θ is a constant throughout any problem, that it represents an unknown value, and that observations and functions of observations (called estimates) are the only variables that occur. We shall use the symbols x_i, $i = 1, 2, \cdots, n$, to represent observed values of the variable x of the universe, and the symbol F to represent a given function of the observations x_i.

If we choose to consider a given function F as an estimate of θ, we are then interested in the error $F - \theta$. This quantity differs from the so-called residual of least square theory, since we are here interested in the difference between computed and true values, rather than in the difference between observed and computed values. To avoid any possible confusion we shall refer to $F - \theta$ as the *error*. Over the set of all samples of n observations, x_i, the distribution of the errors $F - \theta$ is expressed by means of the distribution function $f(F)$,

which may be computed from the known distribution function of the universe. We shall assume that the function $f(F)$ has been normalized, so that $\int_\alpha^\beta f(F)\,dF = 1$, where the interval from α to β includes all possible values of F. The integral $I = \int_\alpha^\beta (F - \theta)^2 f(F)\,dF$, associated with a given estimate F, may be thought of as the average square error over the set of all samples.

In this notation we shall state a criterion for the judgment of estimates in either of the two following forms:

DEFINITION 1. *Let f_1 be the distribution function of F_1, and f_2 that of F_2. The estimate F_1 of θ will be judged better than the estimate F_2 if*

$$\int_\alpha^\beta (x - \theta)^2 f_1(x)\,dx < \int_\alpha^\beta (x - \theta)^2 f_2(x)\,dx.$$

DEFINITION 2. *From a given class of functions, of which F is a member, F will be called the best estimate if*

$$I = \int_\alpha^\beta (F - \theta)^2 f(F)\,dF \tag{1}$$

is less than the corresponding integral for all other functions of the class.

It is to be observed that the integral I is a function of the quantities θ and f. From this is seen at once the distinction between the present problem of minimizing the average square error and the similar problem of finding that point around which the mean square value of the deviations of a variable is a minimum. In the problem under consideration we wish to find the function F, or more precisely its distribution function $f(F)$, for which I takes its minimum with a fixed value of θ. In the alternative problem we have a given distribution f and we wish to find the minimum of I with respect to θ.

A second observation to be made is that the integral I can not be usefully minimized in the sense of the general conditions of the calculus of variations. The problem would be of the isoperimetric variety, with the side condition $\int_\alpha^\beta f(x)\,dx = 1$. A solution might be expressed as the limit, as a approaches zero, of functions $f(x)$ with proper continuity conditions, such that

$$f(x) \begin{cases} = 0 \text{ when } |x - \theta| \geqq a, \\ > 0 \text{ when } |x - \theta| < a, \text{ and } \int_{\theta-a}^{\theta+a} f(x)\,dx = 1. \end{cases}$$

Such a solution would be meaningless in practical statistical theory. Solutions are to be expected, therefore, only in those cases where the class of functions, from which F is to be selected, is sufficiently restricted.

The two following examples illustrate both restrictions and possible application of the theory.

As a first example let us consider the problem of finding an estimate F of the mean, $\bar{x}$, of a normal universe. The mean of a distribution is a symmetric linear function of the variates of the distribution. For the class of functions from which to select an estimate F of $\bar{x}$, let us take the class of all symmetric homogeneous linear functions of the observations x_i. Let

$$F = a(x_1 + x_2 + \cdots + x_n). \tag{2}$$

We wish to find the value of a, if any, for which I is a minimum.

F is the sum of n normally distributed independent variables, ax_i, each with standard deviation $a\sigma$. F, therefore, has a distribution function

$$f = C \cdot \exp\left(\frac{-(F - an\bar{x})^2}{2a^2 n\sigma^2}\right),$$

where C is so chosen that $\int_{-\infty}^{\infty} f\,dF = 1$. A discussion of general distribution functions may be found in Dunham Jackson's article, "Theory of Small Samples," in the *American Mathematical Monthly*, Volume XLII, 1935. In this case it can be shown without particular difficulty that

$$\begin{aligned} I &= C \int_{-\infty}^{\infty} (F - \bar{x})^2 \cdot \exp\left(\frac{-(F - an\bar{x})^2}{2a^2 n\sigma^2}\right) dF \\ &= a^2 n\sigma^2 + \bar{x}^2(an - 1)^2. \end{aligned}$$

To determine the minimum of I with respect to a, we set

$$\frac{\partial I}{\partial a} = 2\,an\sigma^2 + 2\bar{x}^2(an - 1)n = 0,$$

and obtain

$$\begin{aligned} a &= \frac{\bar{x}^2}{n\bar{x}^2 + \sigma^2} = \frac{1}{n}\,\frac{1}{1 + \sigma^2/n\bar{x}^2} \\ &= \frac{1}{n}\left(1 - \frac{\sigma^2}{n\bar{x}^2} + \cdots\right). \end{aligned} \tag{3}$$

It is seen that for even such a simple example as the estimation of the mean there is no estimate of the form of equation (2), with a independent of the parameter to be estimated, for which I takes its minimum value.

For a distribution in which $\bar{x} \neq 0$, and $\sigma^2/n\bar{x}^2$ is small, a is given as a first approximation by $1/n$. The function F is merely the mean of the sample observations. If $\bar{x} = 0$, the required solution is $a = 0$, and there is no best least square estimate of the type of equation (2).

In the case where $\sigma^2/\bar{x}^2$ is not small, as is apt to be the case when $\bar{x}$ is near zero, the determination of a desirable estimate by least squares requires a knowledge of the ratio $\sigma^2/\bar{x}^2$, which may perhaps be judged approximately in a special

problem. If this value is assumed known, the required value of a may be found most easily by rewriting equation (3) in the form

$$a = \frac{1}{n + \sigma^2/\bar{x}^2}. \tag{4}$$

The second example to be considered is the determination of an estimate of σ^2 of a normal universe. A comparison with the definition of σ^2 suggests the use of a function F given by the equation

$$F = a\{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\}, \tag{5}$$

where $\bar{x}$ is the mean of the n observations. The value of a is, of course, to be determined by minimizing the integral I.

F is the sum of the squares of n normally distributed but not independent variables. It may be shown, however, (Jackson, *loc. cit.*) to be expressible as the sum of the squares of $n-1$ independent normally distributed variables, each with standard deviation $\sqrt{a}\sigma$. The distribution function for F takes the form

$$f(F) = C\,(F)^{(n-3)/2}e^{-F/2a\sigma^2}, \tag{6}$$

F taking only positive values, and C is again chosen to normalize $f(F)$. The integral I may be written

$$I = C\int_0^\infty (F - \sigma^2)^2(F)^{(n-3)/2}e^{-F/2a\sigma^2}\,dF.$$

The integration is most easily accomplished by replacing F by u^2, and in terms of u

$$I = C'\int_0^\infty (u^2 - \sigma^2)^2u^{n-2}e^{-u^2/2a\sigma^2}\,du.$$

The various steps in the integration will differ for even and odd values of n, but in each case the final result is the same. It is found that

$$I = \sigma^4\{a^2(n^2 - 1) - 2a(n - 1) + 1\}. \tag{7}$$

The value of a which minimizes I is determined from the relation

$$\frac{\partial I}{\partial a} = \sigma^4\{2a(n^2 - 1) - 2(n - 1)\} = 0.$$

Dividing by $(n-1)$, which is not zero in a sample of two or more observations, we obtain

$$a = \frac{1}{n + 1}. \tag{8}$$

In contrast to the previous example we have here an absolute minimum of I with respect to all estimates of the type of equation (5). The best least square estimate of this type is, therefore,

$$F = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n + 1}. \tag{9}$$

Pennsylvania State College,
State College, Pa.

THE TEACHING OF STATISTICS[1]

By Harold Hotelling

The very great increase in the teaching of statistics since the First World War has been associated on one hand with the development of statistical theory. This important series of discoveries has made available more and more powerful and accurate statistical methods, and has also acquired an intellectual interest of its own as embodying the modern version of the most important part of inductive logic and as providing scope for mathematical and logical ingenuity of high order. The increased teaching of statistics has also been associated with the rapidly growing applications of statistics in innumerable fields, made possible by the development of the theory, by the availability of persons having some knowledge of the theory, and by an increasing realization of the possibilities of application. Doubtless most students of statistics enter upon the subject, not for its intrinsic interest, but with the idea of applying statistical methods as a tool to some particular end. This object may be scientific research, or to fulfill a requirement for a degree, but is often connected with some purely practical pursuit offering the ready prospect of a remunerative job. But it would be a mistake to ignore those whose interest is more purely intellectual, who desire an insight into the peculiar problems of probable inference and the structure of empirical knowledge, who wish to get a fundamental acquaintance with one of the most fundamental of subjects, to see and understand fully the mathematical derivations underlying so much practical and scientific activity, and perhaps to make their own contributions.

Of the magnitude of the demand for statisticians there can be no doubt. The realization of what statistical methods can do in a multitude of fields has gradually led the administrators of government agencies, directors of scientific organizations and research institutes, and business men, to employ rapidly increasing numbers of persons with some knowledge of statistical methods, and to accord an unusual degree of recognition and promotion in many such cases. The uses of statistical methods, and especially of sampling theory, are so varied that it is scarcely possible in a brief space to give any sort of survey of them. They enter, in one form or another, into the research work of the physicist, the chemist, the astronomer, the biologist, the psychologist, the anthropologist, the medical investigator, the economist, and the sociologist. Meteorology, which has lately acquired greatly increased importance, both civil and military, is with its masses of numerical observations very much a statistical matter. The engineer needs modern statistical methods both in the physical and in the

[1] Address at the meeting of the Institute of Mathematical Statistics at Hanover, N. H., September 10, 1940.

economic aspects of his plans. The work of W. A. Shewhart has made clear the central importance of sampling theory in the economic control of quality of manufactured articles. Business men who use sampling surveys to test the markets for their products and the effectiveness of their advertising, who employ statisticians to make up index numbers and forecasts of business conditions, and whose manufacturing costs and quality are controlled with the help of recently devised statistical methods, are finding more and more uses for statisticians. Indeed, it seems as if the exploitation of the business and manufacturing possibilities of statistical methods has only begun, and that limitless further fields are coming into view. Insurance has of course always been essentially dependent on statistics.

But the most rapidly growing large class of positions for statisticians is at present in governmental activities. For some facts regarding the employment of statisticians by the federal government I am indebted to Dr. J. M. Thompson. It appears that it has about one hundred agencies using statistics, with almost eight hundred positions broadly classified as statistical or mathematical, in addition to more than six thousand generally classified as economists. The title "economist" covers many types of work, but much of it is largely statistical. The nature of the government's statistical work is varied and extensive. It includes such work as forecasting revenue from taxes, prices and production of agricultural commodities, general demand conditions, and weather. Some of the work consists in analyzing the effects of various taxes on other programs. In connection with proposed legislation, statisticians serving the lawmakers often attempt to outline the probable results of the legislation, as well as to assist in setting up definite formulae for carrying out the general policies aimed at in Acts of Congress. Administrators as well as lawmakers require statistical activities of a high order, exemplified in the Bureau of the Census, the Bureau of Agricultural Economics, and others. The scientific activities of the government, the work of the War Department, and many others that do not at first sight appear at all statistical, require the services of mathematical statisticians of high order. Even the judicial activities call for statistical theory of some of the most recently discovered kinds, as for instance in the investigation recently made of parole procedures. Cities and states, school and port authorities, employ numerous statisticians for other and widely diverse purposes.

The growing need, demand and opportunity have confronted the educational system of the country with a series of problems regarding the teaching of statistics. Should statistics be taught in the department of agriculture, anthropology, astronomy, biology, business, economics, education, engineering, medicine, physics, political science, psychology, or sociology, or in all these departments? Should its teaching be entrusted to the department of mathematics, or to a separate department of statistics, and in either of these cases should other departments be prohibited from offering duplicating courses in statistics, as they are often inclined to do? To what students, and at what stage of their advancement, should a course in statistics be administered?

Should there be mathematical or other prerequisites? How much of an investment in a statistical laboratory is warranted? Should courses be primarily theoretical and mathematical, or should they be made as practical as possible, equipping the student in the shortest possible time for a job as statistician, or for statistical work in the field with which a particular department is concerned? What about degrees in statistics? Eclipsing all these in importance, though it seems to have received too little of the attention of college and university administrative officers is the question, What sort of persons should be appointed to teach statistics?

To pressing practical problems answers are sure to be given either by considered policy or by processes of historical evolution. The latter are the more prominent in explaining the statistical teaching we have had. A synoptic picture of the origins, not many decades ago, of a good deal of it would perhaps be something like this. A university Department of X, where X stands for economics, psychology, or any one of numerous other fields, begins to note toward the end of the pre-statistical era that some of the outstanding work in its field involves statistics. The quantity and importance of such work are observed to increase, while at the same time its intelligibility seems to diminish. Evidently students turned out with degrees in the field of X who do not know something about statistics are going to be handicapped, and are not likely to reflect credit on Alma Mater. The department therefore resolves that its students must acquire at least an elementary knowledge of the fundamentals of statistics. To implement this principle, it perhaps inserts some acquaintance with statistics among the requirements for a degree. This situation naturally calls for the introduction of a course in statistics. Accordingly the head of the Department of X, in preparing the next Announcement of Courses, writes:

> "X 82. Elements of Statistics. An elementary but thorough course designed to acquaint students of X with the fundamental concepts of statistics and their applications in the field of X. The viewpoint will be practical throughout. Second semester, MWF at 10.
>
> "Instructor to be announced."

The problem now arises of finding someone to teach the new course. The few well-known statisticians in the country have positions elsewhere from which it would be impossible to dislodge them with the bait to be offered; for though the department wishes to have statistics taught as an auxiliary to the study of X, it feels that there must be no question of the tail wagging the dog, and that economy is appropriate in this connection. The members of the department of professorial rank do not respond favorably to the suggestion that they should themselves undertake to teach the new and unfamiliar course. But every university department has a bright graduate student whose placement is an immediate problem. Young Jones has already demonstrated a quantitative turn of mind in the course on Money and Banking, or in the Ph.D. thesis on which

he has already made substantial progress, dealing with The Proportion of Public School Yard Areas Surfaced with Gravel. He may even recall having had a high-school course in trigonometry. His personality is all that might be desired. He is a white, Protestant, native-born American. And so the "Instructor to be announced" materializes as Jones.

This earnest young scholar now finds that, in addition to completing his thesis, he must look up the literature of statistics and prepare a course in the subject. His attention is directed by older members of the department to some of the research papers in the field of X involving statistics. He pursues "statistics" through the library card catalog and the encyclopedias. He reads about census and vital statistics, price statistics, statistical mechanics. Perhaps he encounters probable errors. Eventually he learns that Karl Pearson is the great man of statistics, and that Biometrika is the central source of information. Unfortunately most of the papers in Biometrika and of Pearson's writings, while not lacking in vigor, trail off into mathematical discourse of a kind with which young Jones feels ill at ease. What he wants is a textbook, couched in simple language and omitting all mathematics, to make the subject clear to a beginner. Perhaps he finds the impressive books of Yule and Bowley, but decides that they are too abstruse. Elderton's "Frequency Curves and Correlation" is far too mathematical. Jones decides that a simple book on statistics must be written, and that he will do it if he can ever succeed in mastering the subject. In the meantime, he contents himself perforce with the less mathematical writings of Karl Pearson, with applied examples in the field of X, and with such nonmathematical textbooks as may have been written by other young men who have earlier trod the same path as that on which Jones is now beginning. Somehow or other he gets the class through the course. After doing this two or three times, Jones is an experienced teacher of statistics, and his services are much in demand. His course expands, takes on a settled form, and after a while crystallizes into a textbook. At the same time he may be getting out some research, consisting of studies in the field of X in which statistical methods play a part. His promotion is rapid. He becomes a Professor of Statistics, and perhaps an officer in a national association. His textbook has a large sale, and is used as a source by other young men writing textbooks on statistics.

The textbooks written in this way form an interesting literary cycle. Measures of "central tendency" and of dispersion are introduced, and the use of one as against another of these measures is debated on every ground except the criterion that modern research has shown to be the important one, the sampling stability. Sampling considerations, indeed, get little attention. The urge to simplify by leaving out the more difficult parts of the subject, and especially the mathematical parts, is accompanied by pride in the great number of examples drawn from real life, that is, actual data that have been collected.

But the most fascinating feature of this literary cycle is the opportunity it offers for research by the standard methods of literary investigation, tracing the

influence of one author upon another through parallelism of passages, and so forth. This study is facilitated by the accumulation of errors with repeated copying. One outstanding example is in certain formulae connected with the rank correlation coefficient, derived originally by Karl Pearson in 1907 and copied from textbook to textbook without adequate checking back. As one error after another was introduced in this process, the formulae presented to students (and apparently made the basis of class exercises involving numerical substitution) became less and less like Pearson's original equations. Incidentally, in trying to check this original work of Pearson's, recent investigation has raised the suspicion that it is erroneous; at any rate, he does not give a fully adequate argument. Thus it may be that the errors in copying, which are so useful in examining the history of statistics, never did any harm. The formulae in which the students were drilled may have been no worse than they would have been if all the copying had been done with more care.

While this process has been going on in the Department of X, the Y and Z Departments have likewise evolved the teaching of statistics. There is some interchange of ideas between the various statisticians on the campus, and there is a catholicity in the copying of textbooks. But by and large, statistics is regarded in the Economics Department as a branch of economics, in the Psychology Department as a part of psychology, and so forth. The astronomer is inclined to resent the suggestion that his students should be called upon to study their least squares with anyone but an astronomer. Medical and biological investigators suspect Economics and Psychology of charlatanry, and do not look with favor on the idea of turning their own students over to such departments for instruction in statistics. Most unthinkable of all would be putting the Department of Education in charge of an essential part of the training of scientific students. Thus the courses multiply.

The fact that it is essentially the same fundamental subject that is being taught under various names and with various kinds of notation in different departments is often concealed by including the teaching of statistical theory in a course whose title and prospectus are more suggestive of applications. A case in point is that of an economist of my acquaintance, not primarily engaged in teaching, who some years ago was invited to give a course in Price Forecasting in the Economics Department of a leading university. He carefully prepared a series of lectures on this subject, which had been the center of some extended research he had conducted. A large class enrolled for the course. But soon after beginning his series of lectures the economist noticed that the class was growing restive. Upon inquiring what was amiss, he learned that his discourse was unintelligible to many of them because he was using technical statistical terms and concepts with which they were not familiar. He thereupon undertook to use simpler language, and when this did not suffice to convey his meaning, to explain the statistical notions involved in his work on price forecasting. More and more his lectures came to deal with the elements of statistics, and less and less with price forecasting. At the end of the term he felt that he had

given the students some elementary knowledge of statistical theory, for which they had not enrolled and for which he did not feel particularly well qualified, but had taught them virtually nothing about price forecasting. When the invitation was repeated the next year, the economist suggested imposing a course in statistics as a prerequisite for the course in Price Forecasting. This however was vetoed by the head of the Economics Department, who did not believe in prerequisites. The Price Forecasting course was not repeated.

This incident illustrates the evolution of a good deal of statistical teaching. At the beginning, the idea is to teach some application, but the teacher soon finds himself engaged at much more length than expected with the fundamentals of statistical theory and methods. In this way it has come about that a large number of persons are teaching theoretical statistics who initially had no intention of doing so, but were concerned with particular applications. The teaching of statistical theory has been undertaken belatedly and inexpertly because it was necessary to a discussion of some application originally in view. Thus it happens that a good deal of teaching of statistics, even of mathematical statistics, masquerades as something else.

The obvious inefficiency of overlapping and duplicating courses given independently in numerous departments by persons who are not really specialists in the subject leads to the suggestion that the whole matter be taken over by the Department of Mathematics. This is a promising solution, but it is doomed to failure if, as has sometimes happened, it means that the teaching of statistics is put under the jurisdiction of those who have no real interest in it. Moreover the teaching of statistics cannot be done appreciably better by mathematicians ignorant of the subject than by psychologists or agricultural experimenters ignorant of the subject. The latter indeed have a certain advantage in that the problems seem more real and definite to them; they can sense the difference between the important and the unimportant questions, even if they cannot express the questions in clear mathematical language, and can sometimes arrive intuitively at a correct result that leaves the mathematician puzzled. Also, they can understand more readily than can the mathematician the examples, drawn largely from biological material, which play so important a part in some of the leading expository work on statistics, such as R. A. Fisher's *Statistical Methods for Research Workers.* The pure mathematician has only one advantage over the non-mathematical worker in empirical fields: he is able to set about reading the serious literature of statistical theory. But he must still find this scattered literature, sort it out from a mass of rubbish, fallacies, and false starts, and trace it back historically until he can understand the notation and the presuppositions. He must also contend with the fact that a good deal that is important in statistics is still a matter of oral tradition, and some consists of laboratory techniques. In short, he needs a teacher before he himself sets out to teach the subject. When a Department of Mathematics calls in a young Ph.D., however brilliant, to teach statistics as a part or all of his program, the best thing it can do, if he has not already had a training in modern statistics, is to

give him a furlough for a year or two to enable him to go where he can acquire such a training.

Qualifications of a good teacher of statistics include, first and foremost, a thorough knowledge of the subject. This statement seems trivial, but it has been ignored in such a way as to bring about the present unfortunate situation. Mathematicians and others, who deplore the tendency of Schools of Education to turn loose on the world teachers who have not specialized in the subjects they are to teach, would do well to consider their own tendency to entrust the teaching of statistics to persons who not only have not specialized in the subject, but have no sound knowledge of it whatever. A knowledge of theoretical statistics is not easy to obtain. There is no comprehensive treatise on the subject, starting from first principles, and proceeding by sound deductions and well-chosen definitions to the methods that need to be used in practice. (I have been trying for years to write such a treatise, but it has turned out to be a bigger task than at first appeared. This is partly because some things formerly thought to have been proved turn out, on critical examination, not to be sound, and much new research has been necessary.) The literature is scattered through journals pertaining primarily to many kinds of applications, and it is only in recent years that any large proportion of the current contributions to statistical theory and methods have been gathered into a few periodicals devoted to statistical theory. On the other hand, the seeker after truth regarding statistical theory must make his way through or around an enormous amount of trash and downright error. The great accumulation of published writings on statistical theory and methods by authors who have not sufficiently studied the subject is even more dangerous than the classroom teaching by the same people.

A good teacher of statistics needs of course a mathematical background, including at least an acquaintance with the theory of functions and n-dimensional euclidean geometry. A good deal of additional algebra and analysis are likely to be helpful, as well as some differential geometry. But no amount of such mathematics constitutes by itself any approach to sufficiency in the qualifications of a teacher of statistics. The most essential thing is that the man shall know the theory of statistics itself thoroughly from the ground up, including the mathematical derivations of proper methods and a clear knowledge of how to apply them in various empirical fields. In addition to the pure mathematics and the knowledge of statistical theory, a competent statistician or teacher of statistics needs a really intimate acquaintance with the problems of one or more empirical subjects in which statistical methods are applied. This is quite important. Sometimes excellent mathematicians have wasted time and misled students through failure to get that feeling for applications that is necessary for proper statistical work.

The theory of statistics has been making advances so rapid and so fundamental that some of the first things that need to be said in an elementary course, even for prospective practical statisticians, are affected by some of the most recent researches. So elementary a question as "What definition is it wise to give to

the term 'standard deviation'?", which must be faced by every teacher of Statistics 1, requires for an intelligent answer a rather thorough understanding of modern sampling theory and techniques. The answer, it now seems, is *not* the definition given in most textbooks. In the selection of a statistic to represent a parameter, for example in fitting frequency curves or in linkage estimation in genetics, the fundamental consideration is connected with the sampling distribution, as R. A. Fisher showed in founding the modern theory of estimation. This is ignored in most of the current teaching of statistics, with the result that innumerable students are sent out to waste the money and time of their employers by demanding larger samples than are necessary for the purposes in view, wasting costly information by calculating inefficient statistics and using tests that are not the most powerful. On the other hand, students of statistics who are taught rule-of-thumb methods without their derivations are never quite conscious of the exact limitations and assumptions involved, and may make unwarranted inferences from samples that are too small or in some way violate the conditions underlying the derivations of the formulae.

A good teacher of statistics must be thoroughly familiar with these recent advances. He must examine very critically textbook statements unsupported by full proofs. Even though the students are not capable of following the complete mathematical argument—indeed, especially if the students are not to examine it—the instructor needs to give it a critical study. The custom of omitting proofs, which would not be tolerated in pure mathematics beyond a very limited extent, is common in the teaching of statistics, and is excused on the ground that the students do not know enough mathematics to understand the proofs. Perhaps in some cases a better reason is that the teachers, and the authors of the textbooks, do not understand the proofs. In some instances no proofs exist, and in some instances no genuine proofs can exist, because the methods taught are demonstrably wrong. The custom prevalent in the teaching of mathematics of going over each proof carefully in the class is, among other things, a safeguard against infiltration of false propositions. This safeguard is missing from most of the teaching of statistics, and there has been an infiltration of errors. Since it is accepted that a great many students need to learn something about statistical methods without learning enough mathematics to understand the proofs, it follows that the elementary teaching of statistics to these students must, if the perpetuation of gross errors is to be avoided, be in the hands of really competent mathematical statisticians. This is perhaps the greatest reform needed in the teaching of statistics today. Until the *elementary* teaching of statistics is conducted by those with a thorough and critical knowledge of current research in statistical theory, of a sort that seems virtually inseparable from participation in that research, there is likely to be a continuation of the laborious drilling of thousands of students in methods that ought never to be used. Here, of all places, is the great need for participation of research workers in elementary teaching.

Teachers and textbook writers might well abandon the idea of telling what

statistical methods are used, and say instead what methods ought to be used. But before they can do this with confidence they must have a very close acquaintance with the research of the last three decades in statistical theory.

How can an appointing officer know whether a prospective teacher of statistics knows his subject? This question requires no answer peculiar to statistics in distinction from other subjects. Publication of research, constituting a contribution to the particular field, has always been accepted as the best proof. A substantial contribution to fundamental statistical theory, which is to be distinguished from the mere application of known statistical methods to empirical data, is the best indication of the kind of scholarship appropriate to a teacher of statistics.

Participation in research is not novel as a criterion of what constitutes a good teacher of a college or university subject, if the subject is Greek literature, physics, chemistry, biology, or indeed any of those departments that have been long enough established to attain with respect to the organization of their teaching a state approximating equilibrium. The more reputable institutions of higher learning have long maintained the principle, though with occasional violations in practice, that the Ph.D. degree or its equivalent, representing among other things the completion of a piece of scholarly research, is a minimum condition for a regular faculty appointment. It has usually been maintained also that the Ph.D. thesis should be a new contribution of a strictly scholarly character to the field of the scholar's competence, and not merely a routine application of known methods to an extraneous field. Thus a thesis offered for the Ph.D. degree in mathematics would be judged by its contribution to mathematics, rather than to physics or accounting. Moreover the regard in which universities have held members of their faculties has been intimately connected with their output of scholarly research. Other criteria of excellence have not been ignored, but research has been recognized in a fairly consistent manner. Some say that there has been an over-emphasis on research, and that more attention ought to be given to other qualities related to teaching. However this may be, the facts remain that scholarly research is something capable of a reasonably objective evaluation by scholars in the field, that it offers the main hope of fundamental progress, and that familiarity with current research is a necessary, though not sufficient, condition for the most important teaching in institutions of higher learning.

A peculiarity of the teaching of statistics, of which in practice the theory of statistics is an essential even if unacknowledged part, is that a good deal of it has been conducted by persons engaged in research, not of a kind contributing to statistical theory, but consisting of the application of statistical methods and theory to something else. A similar situation would exist if the teaching of mathematics were in the hands of an assortment of various kinds of engineers, or if zoology and botany were taught by practicing physicians. The teaching of mathematics and of elementary biology might perhaps gain in liveliness and concreteness by such arrangements, with the accompanying emphasis on the

particular applications of the fundamental sciences. Moreover the engineer might in the course of such teaching refresh his own knowledge of elementary mathematics, while the physician might gain by renewing his acquaintance with elementary biology. Such arrangements might occasionally be made with profit. But if they were the general rule the advantages of specialization would be lost; the fundamental sciences would not be developed in so well-rounded a manner as they are by specialists in them, while the special skills and knowledge of the physician and engineer could not be utilized to the full in their respective professions. Statistical theory is a big enough thing in itself to absorb the full-time attention of a specialist teaching it, without his going out into applications too freely. Some attention to applications is indeed valuable, and perhaps even indispensable as a stage in the training of a teacher of statistics and as a continuing interest. But particular applications should not dominate the teaching of the fundamental science, any more than particular diseases should dominate the teaching of anatomy and bacteriology to pre-medical students. These subjects are not ordinarily taught by practicing physicians, but by anatomists and bacteriologists respectively.

In medical education the principle has been accepted, after a long struggle, that a medical school should have full-time professors engaged primarily in teaching and research, and that such professors should not treat patients except in cases of unusual interest from the standpoint of the science or art of medicine. An analogous principle would be that an institution offering extensive instruction in statistics should have full-time professors engaged in the teaching of and research in statistical theory and methods, without spending time over applied statistical problems excepting insofar as such problems might present novel features calling for the development of new statistical methods or theoretical extensions having interest going beyond the immediate case. Sometimes the complaint is heard in medical schools that the teaching tends to become too theoretical on account of detachment from clinical practice, and a similar difficulty might conceivably develop in connection with statistics; but in neither case does the trouble seem to be beyond the ability of the personnel involved to cure if they have the right background.

A specialist in statistics on a university faculty has a threefold function. In addition to the usual duties of teaching and research, there is a need for him to advise his colleagues, and other research workers, regarding the statistical methods appropriate to their various investigations. The advisory function is a highly important one for the activities of the university as a whole, and should be taken into consideration in adjusting the teaching load. Probably every university statistician is visited from time to time by earnest research workers, deeply engrossed in their respective specialities, speaking technical jargons unfamiliar to the statistician, and seeking his advice on matters concerning which he has a sinking feeling of lack of comprehension. After some hours of psychoanalyzing his visitor the statistician may be able to ascertain what it is he *really* wants to know, and thereafter either refer him to some standard formula, or

more often, undertake a piece of new mathematical research designed to fit the particular problem, and very possibly having value also for a more extended class of problems. The statistician is then very likely to find himself embarked on a co-operative research venture in a field that is new to him.

To function well in this third, the consultative or co-operative function, he must have an unusually large store of general information. No one stands in greater need than he of that knowledge of "something about everything and everything about something" that was once said to be the goal of a liberal education. In planning the education of statisticians and teachers of statistics these considerations point to a somewhat wider diffusion of studies among various fields than is customary in many institutions, especially in graduate work. The co-operation, and their other work, would also be facilitated if research workers in general were more strongly urged to get a training in mathematical statistics at an early stage in their careers.

The problem of departmental organization is secondary to that of getting men having the requisite qualities of extensive mathematical preparation, a thorough knowledge of modern theoretical statistics, an understanding of some fields at least in which statistical methods can be applied, and the type of inquiring mind sometimes described as a "research outlook." A Department of Mathematics may well handle the fundamental teaching in statistics, provided it has men properly qualified for such teaching. If it does not have such men, its teaching of statistics and its inability to provide the needed statistical advice will inevitably tempt the other departments to set up again their own duplicating courses in what amounts essentially to statistical theory and methods, and to repeat the mistakes of the past.

A separate Department of Statistics, if competently staffed, could very well provide advice for the whole institution as well as conducting elementary instruction in statistical methods and theory, both for students having calculus and for those without it, and should certainly carry on advanced teaching and research in statistical theory and methods. But for efficient functioning of the institution as a whole it should be agreed that the Department of Statistics or the Department of Mathematics should do *all* the elementary instruction in statistics, and that courses in statistics in other departments should be confined to applications of the basic theory. Normally such courses in applied statistics in the other departments should require as a prerequisite one or more of the basic courses in the Department of Statistics, or of Mathematics. The basic course to be required as a prerequisite to others should be the one which itself requires calculus as a prerequisite wherever this is practicable. It is practicable for students of engineering, physics, astronomy, and mathematical economics, since these students must have calculus anyhow. Moreover the value of the sequence consisting of calculus, statistical theory and applied statistics, in this order, is so great that many other students are likely to avail themselves of it when it is once established and the true nature and value of statistics are more widely understood.

Exactly how far a Department of Statistics should go in particular applications would have to be decided anew from time to time by its members in the light of changing conditions and interests. It cannot teach everything that goes by the name of statistics. This problem may be exemplified by the case of population and vital statistics. This is a field with close connections with sociology, biology, medicine and insurance. It is cultivated in conjunction with each of these subjects in various places. Some of its most interesting and important phases make use of quite advanced mathematics, as in the work of A. J. Lotka, and in addition there is extensive use, and more extensive need, of the statistical methods centered around sampling theory which are the appropriate domain of a Department of Statistics. Should the study of population and vital statistics be included in a Department of Statistics? I think not, except as a temporary arrangement, or in a small institution, in spite of the history of the word "statistics," which originated in connection with material of this kind, and in one of its meanings is still applied to it. (My use of the unqualified word "statistics" in this paper is in the sense of theory and methods, not in the sense of statistical facts such as those found by the census.) Medical, biological and sociological considerations are prominent in the problems of vital statistics, and one of these departments might well handle the subject. But the vital statistician, like other research workers, should have acquired in the course of his training an intimate familiarity with the statistical theory and methods which are the appropriate province of a Department of Statistics. He also needs mathematics through integral equations, if he is to understand and extend the contributions of Lotka and Volterra. Students of vital statistics should have had an elementary course in statistical theory in the Department of Statistics, preferably the course requiring calculus.

A course in price statistics should be taught by an economist, presumably in the Department of Economics, but might well require as a prerequisite the same elementary courses in statistical theory and methods as would be required in psychology, medicine and other fields. In addition, there are problems of time series analysis whose treatment calls for a mathematical statistician having some acquaintance with both economic and meteorological data. A course on the treatment of time series might appropriately be included in the Department of Statistics, requiring the general elementary course as a prerequisite, and itself serving as a prerequisite for courses in economic and meteorological statistics.

One of the chief obstacles to efficient organization of teaching is the habit of not prescribing prerequisites outside one's own department. But when once the elementary courses in statistics have become established in the hands of well-equipped specialists in statistical theory and methods, in whose competence general confidence can be reposed, the various departments of application will lose their motive for establishing their own duplicating courses, and will be able to cultivate more intensively their respective specialities.

The detection of biases and the details of practical statistical work vary greatly

from one application to another. These, consequently, are matters for the departments concerned with applications rather than with the fundamentals of statistics, and should not be the chief features of a course in elementary statistical methods and theory. The work of a Department of Statistics should be concerned largely with sampling theory, and should emphasize the unity of statistical methods and theory, regardless of the field of application. It should deal with statistics as a coherent science of inductive inference, of the preparation of observations for inference, and of the planning of investigations so as to yield observations from which inferences can best be made.

The question what mathematical prerequisites should be established for the fundamental course in statistical theory must be answered by a compromise between the ideal and what is expedient at a particular time and place. In Europe a large number of students have had a year of calculus before coming to universities, that is, before reaching the age of eighteen. If a university were willing to restrict its entrants to such students (thus automatically solving the problem of overcrowding) it could give them another year of calculus, mixed perhaps with advanced algebra and geometry, and then in their sophomore year give them a thorough course in elementary statistics and probability, based on calculus. These students would then be ready to tackle advanced statistics in the third year in a really effective way. If the teaching of economic theory, physics, chemistry and astronomy were geared to this program in such a way as to make real use of the calculus, the work in these subjects could be made far more efficient, in the sense that more material could be covered effectively in the allotted time, or an equivalent amount of material in less time. If, in addition, all the many departments in which statistical methods and theory are used required these statistical courses as prerequisites, and actually used the materials of these courses in their work, there would be a further huge gain in efficiency. The baccalaureate degree of such an institution would represent a far more thorough knowledge, and command of the tools of research, than is possible without an arrangement putting in this way the fundamentals first.

Institutions unwilling to undertake such a drastic improvement must face more or less delay and inadequacy in the acquisition by their students of the fundamentals of mathematics and of statistics. A division of the students into groups according to mathematical ability ought to be undertaken, and followed by a corresponding division of the elementary statistics course. Students having high mathematical ability could begin the study of statistics after completing calculus, and could look forward to rising ultimately to greater heights in pursuits involving mathematical or statistical knowledge than those of lesser mathematical talents. For these latter there would still be the possibility of acquiring, even without calculus, useful statistical tools; but it is essential that this should be done under the guidance of instructors thoroughly familiar with the mathematics of statistics. The task of leading the blind must not be turned over to the blind. Students possessing the ability to master the calculus should

be encouraged to begin the study of statistics with the course having calculus as a prerequisite, and should not be put into the necessarily slower group not having the calculus. I believe that these elementary courses should begin with the theory of probability, but should go on to the chief distribution functions used in practice, and should include applied problems and work on calculating machines.

Putting a sound program of statistical teaching into effect will take time, partly because of the scarcity of suitable teachers of statistics. Nevertheless, the process is well under way, and the prospects are good for substantial improvements in the teaching of statistics. A body of able young research men possessing the requisite knowledge of statistical fundamentals is now in existence and is growing. Some of the recent textbooks represent striking improvements. The Institute of Mathematical Statistics itself, with the *Annals of Mathematical Statistics*, is perhaps the best evidence of a changed view making for better things.

Columbia University,
New York, N. Y.

DISCUSSION OF PROFESSOR HOTELLING'S PAPER

By W. Edwards Deming

It is a pleasure to endorse Professor Hotelling's recommendations; in fact we have been following them pretty closely in the courses in the Graduate School of the Department of Agriculture. As a matter of fact, he has indirectly played an influential part in building up this set of courses, because some of our best instructors are his former students.

Listening to Professor Hotelling's paper, I was thinking of the possibility that some of his recommendations might be misunderstood. I take it that they are not supposed to embody all that there is in the teaching of statistics, because there are many other neglected phases that ought to be stressed. In the Bureau of the Census the population division alone has augmented its force by approximately 3500 statistical clerks during the past six months. They come from diverse schools and it has been interesting to observe how many of them have the idea that all the problems of sampling and inference from data can be solved by what are commonly known as modern statistical techniques—correlation coefficients, rank correlation coefficients, chi-square, analysis of variance, confidence limits, and the like. Most of them are shocked to learn that many of the so-called modern "theories of estimation" are not theories of estimation at all, but are rather theories of distribution and are a disappointment to one who is faced with the necessity of making a prediction from his data, i.e., of basing

some critical course of action on them. The conviction that such devices as confidence limits and Student's t provide a basis for action regardless of the size of the sample whence they were computed, even under conditions of statistical control, is too common a fallacy. On the other hand, many simple but worthy devices are neglected. A histogram, for instance, can be a genuine tool of prediction if it is built up layer by layer in different legends so as to distinguish the different sources whence the data are derived. The modern student, and too often his teacher, overlook the fact that such a simple thing as a scatter diagram is a more important tool of prediction than the correlation coefficient, especially if the points are labeled so as to distinguish the different sources of the data. Most students do not realize that for purposes of prediction the consistency or lack of it between many small samples may be much more valuable than any probability calculations that can be made from them or from the entire lot. Students are not usually admonished against grouping data from heterogeneous sources. Of those that are not guilty of indiscriminate grouping, many are inclined to rely on statistical tests for distinguishing heterogeneity, rather than on a careful consideration of the sources of the data. Too little attention is given to the need for statistical control, or to put it more pertinently, since statistical control (randomness) is so rarely found, too little attention is given to the interpretation of data that arise from conditions not in statistical control.

Nevertheless, the fundamentals of probability and sampling theory, and the mathematics of the distribution functions, though by themselves they do not qualify anyone for high-grade statistical work, are ultimately essential for proficiency in statistics. Since they are seldom learned away from the university they are properly made the main theme of teaching. The university is the place to learn the studies that are so difficult to get outside of it.

Above all, a statistician must be a scientist. The skepticism of many first class scientists of today for modern statistical methods should be a challenge to statistical teaching. A scientist does not neglect any pertinent information, yet students of statistics are often taught to do just the opposite of this, and are accused of being old-fashioned for daring to think of combining experience with the new information provided by a sample, even if it is a pitifully small one. Statisticians must be trained to do more than to feed numbers into the mill and grind out probabilities; they must look carefully at the data, and take account of the conditions under which each observation arises. It is my feeling that the chief duty of a statistician is to help design experiments in such a way that they provide the maximum knowledge for purposes of prediction; another is to compile data with the same object in view; and still a third function is to help bring about some changes in the source of the data. Scientific data are not taken merely for inventory purposes. There is no use taking data if you don't intend to do something about the sources whence they arise.

Bureau of the Census,
Washington

RESOLUTIONS ON THE TEACHING OF STATISTICS

The Institute of Mathematical Statistics at its business meeting on September 11, 1940 at Dartmouth College adopted the following resolutions regarding the teaching of statistics. The resolutions were drawn up by a committee appointed by the President, and consisting of Burton H. Camp, W. Edwards Deming, Harold Hotelling, and Jerzy Neyman.

1. If the teaching of statistical theory and methods is to be satisfactory, it should be in the hands of persons who have made comprehensive studies of the mathematical theory of statistics, and who have been in active contact with applications in one or more fields.

2. The judgment of the adequacy of a teacher's knowledge of statistical theory must rest initially on his published contributions to statistical theory, in contrast with mere applications, in a manner analogous to that long accepted in other university subjects.

3. These ideas are expressed in detail in the paper *The teaching of statistics*, by Professor Harold Hotelling, and the Institute decides to give both the resolution and the paper as wide a circulation as possible.

REPORT OF THE HANOVER MEETING OF THE INSTITUTE

The sixth meeting of the Institute of Mathematical Statistics was held at Dartmouth College, Hanover, New Hampshire, Tuesday to Thursday, September 10 to 12, 1940, in conjunction with meetings of the American Mathematical Society and of the Mathematical Association of America. The following forty-two members of the Institute attended the meeting:

H. E. Arnold, Felix Bernstein, G. W. Brown, J. H. Bushey, B. H. Camp, A. T. Craig, A. R. Crathorne, J. H. Curtiss, J. F. Daly, W. E. Deming, J. L. Doob, Churchill Eisenhart, M. L. Elveback, C. H. Fischer, M. M. Flood, R. M. Foster, T. C. Fry, H. P. Geiringer, Robert Henderson, E. H. C. Hildebrandt, G. M. Hopper, Harold Hotelling, E. V. Huntington, M. H. Ingraham, Dunham Jackson, W. L. Kichline, L. F. Knudsen, B. A. Lengyel, W. G. Madow, J. W. Mauchly, Richard von Mises, E. B. Mode, Jerzy Neyman, P. S. Olmstead, Oystein Ore, M. M. Sandomire, L. W. Shaw, F. F. Stephan, A. G. Swanson, Abraham Wald, S. S. Wilks, Jacob Wolfowitz.

The meeting of the Institute consisted of four sessions. At the first session, which was held on Tuesday morning, Professor Harold Hotelling of Columbia University delivered an address on *The Teaching of Statistics*. This address was followed by considerable discussion on the various aspects of the teaching of statistics.[1] Preceding Professor Hotelling's address a short paper on an *Empirical Comparison of the "Smooth" test for goodness of fit with Pearson's Chi-Square test* was presented by Professor J. Neyman of the University of California.

Following Professor Hotelling's address a business meeting of the Institute was held. At this time resolutions on the teaching of statistics were approved (see p. 472). The President reported that a War Preparedness Committee had been appointed in the summer to study the matter of the Institute's participation in the national defense program.[2] The Chairman of this Committee submitted a preliminary report which met the approval of the Institute. A plan was approved for completing the report and circularizing it with a minimum of delay.

The matter of the organization of local sections or chapters of the Institute was discussed but no action was taken.

---

[1] Professor Hotelling's address and three resolutions regarding the teaching of Statistics which were adopted by the Institute at a business meeting following the address are published in the present issue of the *Annals of Mathematical Statistics*, pp. 457–472.

[2] The membership of the Committee is as follows:
Professor Churchill Eisenhart (Chairman), University of Wisconsin.
Professor A. T. Craig, University of Iowa.
Professor E. G. Olds, Carnegie Institute of Technology.
Captain Leslie E. Simon, Aberdeen Proving Ground.
Mr. Ralph E. Wareham, General Electric Company.

On Tuesday afternoon a session on contributed papers in Mathematical Statistics was held jointly with the American Mathematical Society. Professor B. H. Camp of Wesleyan University presided and the following papers were presented:

1. *Contributions to the theory of the representative method of sampling.*
   Dr. W. G. Madow, Department of Agriculture, Washington.
2. *A generalization of the law of large numbers.*
   Dr. Hilda P. Geiringer, Bryn Mawr College.
3. *On the problem of two samples from normal populations with unequal variances.*
   Professor S. S. Wilks, Princeton University.
4. *Experimental determination of the maximum of an empirical function.*
   Professor Harold Hotelling, Columbia University.
5. *Asymptotically shortest confidence intervals.*
   Dr. Abraham Wald, Columbia University.
6. *Reduction of certain composite statistical hypotheses.*
   Dr. G. W. Brown, R. H. Macy and Company, Inc., New York.
7. *Conception of equivalence in the limit of tests and its application to certain λ and χ^2 tests.*
   Professor J. Neyman, University of California.

Abstracts of these papers follow this report.

On Wednesday morning a session was held on *The Theory of Probability* with Dr. T. C. Fry of the Bell Telephone Laboratories, in the chair. The following addresses were given:

1. *On the foundations of probability theory.*
   Professor R. von Mises, Harvard University.
2. *Probability as measure.*
   Professor J. L. Doob, University of Illinois.

This session was followed by an energetic discussion which was continued in an informal afternoon session.

The Thursday morning session was devoted to the *Theory of Statistical Estimation* with Professor Harold Hotelling as Chairman. The following addresses were given:

1. *Estimation by intervals as a classical problem in probability.*
   Professor J. Neyman, The University of California.
2. *Statistical estimation in large samples.* Dr. Joseph F. Daly, The Catholic University of America.

On Monday at 4:15 p.m. a tea was held at the Graduate Club for members of the mathematical organizations and their guests, and on Monday at 8:00 a musical performance was presented. On Tuesday at 7:00 p.m. a joint dinner was held for the mathematical organizations in Thayer Hall. Wednesday afternoon was devoted to an excursion to Franconia Notch.

During the meeting a collection of string models of ruled surfaces was exhibited by Professor Robin Robinson of Dartmouth College and electrical calculation apparatus made from telephone equipment was exhibited by members of the staff of the Bell Telephone Laboratories.

ABSTRACTS OF PAPERS

(Presented on September 10, 1940, at the Hanover meeting of the Institute)

Contributions to the Theory of the Representative Method of Sampling. WILLIAM G. MADOW, Washington, D. C.

The theory of representative sampling may be regarded as a dual sampling process; the first of which consists in the sampling of different random variables and the second of which consists in repeating several times the experiments associated with each of the different random variables. It follows that while the theory of sampling from finite populations without replacement may be required for the first process, the second leads directly into the theory of sampling from infinite populations. There is, however, one difference. Although the usual theory is concerned with the evaluation of fiducial or confidence limits for parameters the theory of sampling is concerned with the evaluation of fiducial or confidence limits for, say, the mean of a sample of N, when n, $(N \geq n)$, of the values are known.

It is thus possible to use the usual theories of estimation in obtaining estimates of the parameters and to allow the effects of subsampling process to show themselves in the different values of the fiducial limits. It is shown that the limits obtained are almost identical with those obtained by the theory of sampling from a finite population. Distributions of the statistics used in these limits are derived.

Besides these results, the theory is extended to the theory of sampling vectors, and conditions are stated under which the "best" allocation of the number in a sample among several strata is proportional to the kth roots of the generalized variance of a random vector having k components.

A Generalization of the Law of Large Numbers. HILDA GEIRINGER, Bryn Mawr.

Let $V_1(x)$, $V_2(x)$, $\cdots$, $V_n(x)$ be n probability distributions which are not supposed to be independent and let $F(x_1, x_2, \cdots, x_n)$ be a "statistical function" of n observations in the sense of v. Mises,—$V_i(x)$ $(i = 1, 2, \cdots n)$ indicating as usual the probability of getting a result $\leqq x$ at the ith observation—. Then *it can be proved* that under fairly general conditions $F(x_1, x_2, \cdots, x_n)$ *converges stochastically toward its "theoretical value"*; or in other words, that under these general conditions *a great class of statistics* $F(x_1, x_2, \cdots, x_n)$ *is "consistent"* in the sense of R. A. Fisher.

Well known particular cases of this theorem result if (*a*) we take for $F(x_1, x_2, \cdots, x_n)$ the *average* $(x_1 + x_2 + \cdots + x_n)/n$ of the n observations, (*b*) we assume that the $V_i(x)$ are *independent* distributions.

On the Problem of Two Samples from Normal Populations with Unequal Variances. S. S. WILKS, Princeton University.

Suppose O_{n_1} and O_{n_2} are samples of n_1 and n_2 elements from normal populations π_1 and π_2 respectively. Let a_1, σ_1^2 and a_2, σ_2^2 be the means and variances of π_1 and π_2 and let O_{n_1} and O_{n_2} have means $\bar{x}_1$ and $\bar{x}_2$ and variances s_1^2 and s_1^2 (unbiased estimates of σ_1^2, σ_2^2) respectively. It is shown that there exists no function (Borel measurable) of $\bar{x}_1$, $\bar{x}_2$, s_1^2, s_2^2, $a_1 - a_2$ independent of σ_1 and σ_2, having its probability law independent of the four population parameters. It is therefore impossible to obtain exact confidence limits

for $a_1 - a_2$ corresponding to a given confidence coefficient. Functions of the four parameters and four statistics are devised from which one can set up confidence limits for $a_1 - a_2$ with associated confidence coefficient inequalities.

Experimental Determination of the Maximum of an Empirical Function. HAROLD HOTELLING, Columbia University.

In physical and economic experimentation to determine the maximum of an unknown function, for example of a monopolist's profit as a function of price, or of the magnetic permeability of an alloy as a function of its composition, the characteristic procedure is to perform experiments with chosen values of the argument x, each of which then yields an observation, subject to error, on the corresponding functional value $y = f(x)$. The values of x need, however, to be chosen on the basis of earlier experiments in order to make the determination efficient. The experimentation properly proceeds, therefore, in successive stages, with the values used at each stage determined with the help of the earlier work. The question what distribution of x as a function of previous results should be used is discussed in this paper on the basis of various hypotheses regarding the function, and further criteria. In particular, a conflict is shown to exist under some conditions between the criterion of minimum sampling variance and that calling for absence of bias.

Asymptotically Shortest Confidence Intervals. ABRAHAM WALD, Columbia University.

Let $f(x, \theta)$ be the probability density function of a variate x involving an unknown parameter θ. Denote by $x_1, \cdots, x_n$ n independent observations on x and let $C_n(\theta)$ be a positive function of θ such that the probability that $\left| \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \sum_{\alpha=1}^{n} \log f(x_\alpha, \theta) \right| \leq C_n(\theta)$ is equal to a constant β under the assumption that θ is the true value of the parameter. Denote by $\theta'(x_1, \cdots, x_n)$ the root in θ of the equation $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \sum_{\alpha} \log f(x_\alpha, \theta) = C_n(\theta)$ and by $\theta''(x_1, \cdots, x_n)$ the root of $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \sum_{\alpha} \log f(x_\alpha, \theta) = -C_n(\theta)$. Under some weak assumptions on $f(x, \theta)$ the interval $\delta_n(x_1, \cdots, x_n) = [\theta'(x_1, \cdots, x_n), \theta''(x_1, \cdots, x_n)]$ is in the limit with $n \to \infty$ a shortest unbiased confidence interval[1] of θ corresponding to the confidence coefficient β. This confidence interval is identical with that given by S. S. Wilks in his paper "Shortest average confidence intervals from large samples," *The Annals of Mathematical Statistics*, Sept. 1938. Wilks has shown that $\delta_n(x_1, \cdots, x_n)$ is asymptotically shortest in the average compared with all confidence intervals computed on the basis of statistics belonging to a certain class C. In the present paper it has been proved that the confidence interval in question is asymptotically shortest compared with *any* arbitrary unbiased confidence interval, without any restriction to a certain class of functions.

Reduction of Certain Composite Statistical Hypotheses. GEORGE W. BROWN, R. H. Macy and Co., New York.

The results obtained make it possible to reduce a large class of composite statistical hypotheses to equivalent simple hypotheses. The fundamental theorem established states essentially that if two distributions give rise, in sampling, to the same distribution of the

[1] For the definition of a shortest unbiased confidence interval see the paper by J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc.* (1937).

set of differences between observations, then one distribution must be a translation of the other, subject to a condition requiring that the characteristic function of one of the distributions be such that any interior intervals of zeros be not too large. The result is established by means of the functional equation $\varphi(t_1)\varphi(t_2)\varphi(-t_1 - t_2) = \psi(t_1)\psi(t_2)\psi(-t_1 - t_2)$ relating the characteristic functions. Similar results are obtained for scale, and combination of location and scale, and the corresponding situations in multivariate distributions. This type of uniqueness theorem permits one to reduce a composite hypothesis involving an unknown location parameter (or scale, or both) to an equivalent simple hypothesis.

Conception of Equivalence in the Limit of Tests and Its Application to Certain λ- and χ^2-Tests. J. NEYMAN, University of California.

Denote by E a system of observable variables and by N the number of independent observations of those variables to be used for testing a certain statistical hypothesis H against a set Ω of admissible simple hypotheses h. Let further $T_1(N)$ and $T_2(N)$ be two different tests of H using the same number N of observations. Consider the probability $P_N(h)$ calculated on any admissible simple hypothesis h, of the two tests, contradicting themselves.

Definition: If, whatever be $h \epsilon \Omega$, the probability $P_N(h)$ tends to zero as N is indefinitely increased, then the two tests are said to be equivalent in the limit.

Consider a number s of series of independent trials and denote by $E_{i1}, E_{i2}, \cdots, E_{im_i}$ all the m_i possible and mutually exclusive outcomes of each of the trials forming the ith series. Let p_{ij} be the probability of E_{ij}, n_i the total number of trials in the ith series, and n_{ij} the number of these which give the outcome E_{ij}.

Suppose that it is desired to test a composite hypothesis H concerning all the probabilities p_{ij} and consisting of the assumption that any one of them is a given linear function of some t independent parameters θ_k, so that

$$p_{ij} = a_{ij0} + a_{ij1}\theta_1 + \cdots + a_{ijt}\theta_t \tag{1}$$

where the coefficients a_{ijk} are known. The main result of the paper is then that the λ-test of the above hypothesis H, tested against the set Ω of alternatives ascribing to the p_{ij} any non-negative values, is equivalent in the limit to the test consisting of rejecting H when the minimum of the expression

$$\chi^2 = \sum_{i=1}^{s} \sum_{j=1}^{m_i} \frac{(n_{ij} - n_i p_{ij})^2}{n_{ij}} \tag{2}$$

calculated with respect to unrestricted variation of the θ's, exceeds the tabled value of χ_ϵ^2 corresponding to the chosen level of significance ϵ and to the number of degrees of freedom $\sum_{i=1}^{s} m_i - s - t$.

It will be noticed that the expression (2) differs from the usual χ^2 in the denominator of each term.

As an example of the application of the test based on (2), consider the case where M varieties of sugar beet are tested for resistance to a certain disease in an experiment arranged in N randomized blocks. Denote by n the number of beets selected at random for inspection from each plot and by n_{ij} the number of those of the ith variety from the plot in the jth block which are found to be infected. Denote further by p_{ij} the proportion of infected beets of the ith variety in the plot in the jth block. The hypothesis that the effects of variety and of block are additive is expressed by $p_{ij} = p + V_i + B_j$ with $\Sigma V_i = \Sigma B_j = 0$. To test this hypothesis we may use (2) which in this particular case reduces itself to

$$\chi^2 = \sum_{i=1}^{M} \sum_{j=1}^{N} w_{ij}(q_{ij} - p - V_i - B_j)^2 \tag{3}$$

with $w_{ij} = n^3/\{n_{ij}(n - n_{ij})\}$, $q_{ij} = n_{ij}/n$. The minimum χ_0^2 of χ^2 is found by solving a set of equations which are linear in p, V_i, B_j and the comparison of χ_0^2 with the tabled value corresponding to $(M - 1)(N - 1)$ degrees of freedom will tell us whether we are likely to be very wrong in assuming additivity or not. In the favorable case we may next proceed similarly to test another hypothesis that there is no differentiation between the varieties, so that $V_1 = V_2 = \cdots = V_M = 0$.

Empirical Comparison of the "Smooth" Test for Goodness of Fit with the Pearson's χ^2 Test. J. NEYMAN, University of California.

In a previous publication[2] the author has deduced a test for goodness of fit, described as the "smooth test" or the ψ^2 test, applicable to cases where the hypothesis tested H is simple. The test is so devised as to be particularly sensitive to departures from H which are "smooth" in the sense explained in detail in the publication quoted. Whether the test so devised does present any advantage over the usual χ^2 test depends on how frequently we meet, in practice, cases where the hypotheses alternative to the one tested are actually smooth.

The present investigation was undertaken with the object of obtaining some information on this point. For that purpose a number of cases described in the literature where there was a question of testing that some observable variable x follows some perfectly specified distribution $p(x)$ were analyzed. Of all such cases, the ones where there were *a priori* theoretical reasons to believe that $p(x)$ could not possibly represent the true distribution of x and, at the most, it could be considered as only an approximation to the true distribution were selected.

It was assumed that the departures from the hypothetical distributions are typical of those that may be met in practice when no definite information as to the actual state of affairs is available. The hypothesis of goodness of fit was tested both by means of the χ^2 and by the fourth order smooth test. Out of the 130 cases studied the two tests were in perfect agreement eight times. Out of the remaining 122 cases the smooth test proved to be more sensitive than the χ^2 in 70 cases and the χ^2 better than the smooth test in 52 cases. We may further compare the tests by counting those cases where one of them detected the falsehood of the hypothesis tested at a given level of significance while the other failed to do so. At the level of significance .05 the χ^2 test rejected the hypothesis tested 13 times, while P_{ψ^2} was $>.05$. The reverse was true in 17 cases. At the level of significance .01 the corresponding figures are 5 and 14, again in favor of the smooth test.

---

[2] J. Neyman, " 'Smooth Test' for Goodness of Fit." *Skandinavisk Aktuarietidskrift*, 1937, pp. 149–199.

REPORT OF THE WAR PREPAREDNESS COMMITTEE OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The generally recognized functions of a *statistician* are the calculation of averages, percentages, and index numbers; the construction of bar graphs and pie diagrams; and the compilation of data in general. His other activities are less widely known. In particular, the recent advances in *mathematical statistics* are known to a relatively small proportion of the persons occupying responsible positions in academic life, in industry, and in government. The *mathematical statistician*, in fact, is concerned chiefly with the interpretation of data through the use of probability theory; his is the science of reasoning from a part to the whole, and of prediction; and to him falls the task of stating the conditions under which such inferences are possible, of devising means of testing whether these conditions are satisfied, and of evaluating the probability that such 'uncertain inferences' are correct in specific instances. Furthermore, it is his responsibility to so plan the lay-out of experiments and the conduct of surveys that the data they yield will contain the maximum information on the points at issue and be amenable to unambiguous statistical interpretation.

Because of the functions which the *mathematical statistician* can perform his services should be of value to the National Defensé Program in the following fields:

I. Quality Control and Specification. The functions of a mathematical statistical nature connected with quality control and specification of articles produced by mass production are:

(1) *Tests of randomness.* These are important because statistical methods of inference are strictly valid only for random samples.

(2) *The use of probability theory in predicting the outcome of future repetitions of an operation which is in a state of statistical control.*[1] The evaluation of the probability that the quality of a piece of product will lie within any previously specified tolerance limits as long as a state of statistical control is maintained, and the development of sampling inspection techniques are examples of this function.

---

[1] A repetitive operation, such as a production process, is said to be in a *state of statistical control* when it produces a sequence of observations which exhibit the property *randomness*. An important aspect of quality control is the improvement of quality which comes as the result of an effort to reduce a manufacturing process to a state of statistical control. Furthermore, when this state of control is attained it is possible to gain a reduction in cost of inspection, a reduction in cost of rejections, a reduction in tolerance limits where quality measurement is indirect, and the attainment of uniform quality even though the inspection test is destructive.

(3) *Representative sampling.* When a repetitive operation such as a production process is not in a state of statistical control, it is not possible to make valid inferences about the quality of a lot from an examination of a sample from the lot unless the sampling process is one of random selection within "strata" in accordance with the principles of representative sampling.

(4) *Analysis of variance.* Reference is made here to the technique whereby the total variability of a product of an operation which is in a state of statistical control can be decomposed into components associated with the various sub-operations involved.

(5) *Correlation methods.* When a direct measurement of quality is extremely costly, it is sometimes advisable to use as an indirect measurement of quality the value of some character less costly to measure which is highly correlated with quality.

(6) *Specification of quality as a variable.* Statistical theory, including tests for randomness, must be taken into account in writing quality specifications if the consumer is to be protected against the vagaries of sampling and the producer safeguarded from the incurring of penalties of an unjust chance.

II. **Sampling Surveys.** The importance of conducting sampling surveys in accordance with the principles of *representative sampling* is well established. It is quite possible that such surveys and partial censuses will be needed in connection with the National Defense Program in order to determine the frequency and location of individuals possessing special traits, e.g. persons capable of withstanding the rigours of dive bombing, or persons possessing types of color blindness which render them valuable as observers who can detect camouflage, etc. The "problem of sizes" connected with Stores and Supplies—see below—may require careful preliminary surveys. Also, surveys may be needed to evaluate the effects of various types of propaganda.

III. **Experimentation of Various Kinds.** The mathematical statistician can be of service in connection with experimentation of various kinds undertaken as a part of the National Defense Program since the following aspects of experimentation are of a mathematical statistical nature:

(1) *Randomization.* Since statistical tests for the existence of differences between samples, of correlation, etc. are strictly valid only for random samples, the operation of randomization is of paramount importance in "the comparison of new designs, new materials or alloys, study of contact phenomena under different conditions, corrosion of materials under different atmospheric conditions, and field trial of equipment, to mention only a few." If randomization is not undertaken, observed differences between designs, for instance, may have arisen from non-random assignable differences in the material presented. Furthermore, the validity of tests for significant differences between the effects of various designs rests upon the condition that the variability observed in the effects of each design be of *random* character and free from trends and non-random shifts in magnitude—i.e. the operation of determining the effects

of each design must be in a state of statistical control, to use a phrase employed in quality control.

(2) *Experimental design.* Without careful attention to the lay-out of an experiment, the data it yields may be difficult and even impossible to interpret. Therefore, the principles of experimental design set forth by R. A. Fisher and his followers are of great importance, as are also the special experimental arrangements which have been devised to cope with many of the more usual difficulties met in practice.

IV. **Personnel Selection.** The allocation of individuals to places where they can be of greatest value in the National Defense Program will undoubtedly require tests for mental and physical traits. Although the development and analysis of such tests is largely in the hands of psychometric groups, the use of methods of multivariate statistical analysis in such work renders this field one in which mathematical statistics ought to play an important role.

It is in the above four fields that there is special need for the training and endowments of the *mathematical statistician.* He can also render valuable assistance in the following fields:

V. **Stores and Supplies.**

(1) *Problem of sizes.* Preliminary surveys are likely to prove useful in ascertaining the relative frequencies of demand for the respective sizes of clothing, etc. in different parts of the country.

(2) *Development of procedures for charting the day to day location and movement of stores and supplies.*

(3) *Problem of replacement of parts and equipment.* In many it is more economical to make replacement at statistically determined times, than to wait for complete failure.

VI. **Transportation and Communication.** Probability theory has shown its usefulness in peace time in handling "traffic" problems that arise in telephone and telegraph communication, electric power distribution, etc. No doubt it will find corresponding application to problems in these fields arising out of the National Defense Program.

VII. **Gunnery and Bombing.** Although there is a need in connection with artillery fire for further development of methods of estimating standard deviations from successive differences in order to minimize the biases arising from slowly changing conditions during the period of firing, the principles of artillery fire are quite firmly established and the relatively new science of bombing is likely to present greater opportunities for the application of the methods of mathematical statistics. For instance, in evaluating bombing techniques there is need of statistical methods in separating the constant biases from the random variability.

VIII. **Meteorology.** The extent to which statistical methods are being employed in meteorology can be seen from an examination of the Monthly Weather Review Supplement No. 39, issued April 1940, and entitled "Reports on Critical Studies of Methods of Long-Range Weather Forecasting." There seems to be excellent opportunity here for the application of methods of multivariate analysis and for the development and uses of methods applicable to serially correlated data. Such work would be of value in National Defense so far as it would enable the forecasting of conditions suitable for launching an attack.

IX. **Medicine.** The National Defense Program will probably require the preparation and storage of hormone substances, toxic compounds, drugs, and other medicinal supplies. Since many such are examined for potency, toxicity, etc. by means of animal assays, there will be considerable opportunity here for the sound application of mathematical statistics in planning and interpreting these bioassays.

In nearly all of the above activities the application of mathematical statistics is likely to encounter two major difficulties:

(1) Obtaining an adequate trial of the methods of mathematical statistics.

(2) Supplying persons to occupy key positions in the application of mathematical statistics in a given field—persons competent in mathematical statistics and who possess a sound background in the field of application.

In some of the above activities, e.g. Quality Control, there will be the further difficulty of

(3) Supplying the vast number of slightly trained workers who will gather the data and perform the analyses.

It is with these difficulties in mind that the Committee recommends that the Institute

(1) Prepare a register of Institute members, stating for each member his background, interests, and experience so far as these relate to mathematical statistics and its applications;[2]

(2) Appoint a committee to handle inquiries concerning personnel qualified to deal with particular projects;

(3) Cooperate to the fullest extent in matters pertaining to quality control and specification with the *Joint Committee for the Development of Statistical Applications in Engineering and Manufacturing*, of which the Institute is a sponsor [3]

---

[2] The preparation of this register should be coordinated with any similar undertaking sponsored by the *National Roster of Scientific and Specialized Personnel*, National Resources Planning Board, Executive Office of the President, Washington, D. C.

[3] We suggest the following as possible undertakings in a cooperative program with the Joint Committee:

(1) Requesting statements regarding the potential contribution to National Defense

(4) Undertake such steps as are feasible which will lead to cooperation with other organizations having interests similar to those of the Institute, e.g. the American Statistical Association, the Psychometric Society, and the Econometric Society.

(5) Establish contact with the National Defense Research Committee headed by Dr. Vannemar Bush and coordinate the Institute's activities with those of this national Committee.

In conclusion, we feel that as an organized group the Institute's primary function in relation to the National Defense Program should be to serve as a reservoir of specialists, experienced in the use of the methods of mathematical statistics, who can direct the use of these methods and be of assistance in the development of new techniques as needed. As a secondary, but equally important function, the Institute is in a position to supervise, and perhaps to undertake through the activities of its individual members, the training in mathematical statistics of the individuals who will be needed in the application of whatever statistical programs of the type noted above are undertaken in connection with the National Defense Program. *It is recommended, therefore, that the Institute's interest in the above activities, and its willingness to be called upon, be adequately publicized*, possibly by sending copies of this report to various members of the Government, such as the Chief Signal Officer and the Coordina-

---

of statistical methods in quality control and specification from men prominent in industry who are familiar with recent developments in quality control. Such individuals would be asked to give, where possible, concrete evidence of the value of such methods in their experience—evidence which would be helpful in securing authoritative acceptance of statistical methods in quality control and specification.

(2) The organization of a syllabus on statistical methods for use in evening courses at various industrial centers. (Captain Simon of our Committee is preparing "An Engineer's Manual of Statistical Methods" which will be issued shortly.)

(3) The preparation of a list of topics for inclusion in university courses.

(4) The preparation of a list of suggested reading on statistical methods in quality control and specification, arranged under such headings as "expository," "methodology," etc.

(5) The arrangement of local meetings and round table discussions at some of the universities in a few large industrial centers. Some well known leader of the locality might serve as chairman. To such a meeting would be invited those men in local industries who were interested in the possibility of applying statistical methods to their problems, and the meeting could be thrown open to discussion after a brief paper outlining the accomplishments of statistical methods of quality control in the speaker's experience and stating the advantages to be gained by employing such methods in the mass production of the War Preparedness Program.

(6) Sponsor the preparation of popular expository articles on quality control for industrial journals, Readers Digest, Scientific American, etc., and other activities designed to popularize the subject and gain authoritative acceptance of statistical methods of quality control.

tor of National Defense Purchases and also to the secretaries of appropriate organizations, such as the American Standards Association, with the request that they advise the Institute of any specific action they feel the Institute should take.

A. T. Craig
E. G. Olds
L. E. Simon
R. E. Wareham
C. Eisenhart, *Chairman.*

MATHEMATICAL REVIEWS

an international journal to abstract and review current mathematical literature, including the *Theory of Probability* and *Theoretical Statistics* and applications

---

Volume 1 is now being published

MICROFILM SERVICE

A microfilm copy of any paper reviewed can be obtained at a nominal cost.

Subscription Rates:

$6.50 per year to members of sponsoring organizations	$13.00 per year to others

Sponsored by the American Mathematical Society, the Mathematical Association of America, and other organizations

Send subscriptions to

MATHEMATICAL REVIEWS

Brown University, Providence, Rhode Island, U. S. A.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

SEPTEMBER, 1940 · $1.50 PER COPY · $6.00 PER ANNUM · VOL. 35 · No. 211

Address inquiries and orders for subscriptions and back numbers to R. L. Funkhouser, Secretary, American Statistical Association, 1626 K Street, N. W., Washington, D. C.

BIOMETRIKA

A Journal for the Statistical Study of Biological Problems

Volume XXXI, Parts III and IV

CONTENTS

I. On generalized analysis of variance. By P. L. Hsu. II. The derivation of the fifth and sixth moments of the distribution of b_2 in samples from a normal population. By C. T. Hsu and D. N. Lawley. III. Testing the homogeneity of a set of variances. By H. O. Hartley. IV. The simultaneous distribution in samples of mean and standard deviation, and of mean and variance. By L. Truksa. V. Certain projective depth and breadth measurements of the facial skeleton in man. By Alette Schreiner. VI. Transposition of the viscera and other reversals of symmetry in monozygotic twins. By E. A. Cockayne. VII. The human remains of the iron age and other periods from Maiden Castle, Dorset. By C. N. Goodman and G. M. Morant. VIII. Homogeneity of results in testing samples from Poisson series. With an application to testing clover seed for dodder. By J. Przyborowski and H. Wilenski. IX. On the method of paired comparisons. By M. G. Kendall and B. Babington Smith. X. The mean and variance of χ^2, when used as a test of homogeneity, when expectations are small. By J. B. S. Haldane. XI. A note on the statistical analysis of sentence-length as a criterion of literary style. By C. B. Williams. XII. Applications of the non-central t-distribution. By N. L. Johnson and B. L. Welch. *Miscellanea:* (i) Note on Stirling's Approximation. By E. V. Huntington. (ii) A note on the interpretation of quasi-sufficiency. By M. S. Bartlett. (iii) The cumulants and moments of the binomial distribution, and the cumulants of χ^2 for a $(n \times 2)$-fold table. By J. B. S. Haldane.

The subscription price, *payable in advance*, is 45s. inland, 54s. export (per volume including postage). Cheques should be drawn to Biometrika and sent to "The Secretary, Biometrika Office, Department of Statistics, University College, London, W.C. 1." All foreign cheques must be in sterling and on a bank having a London agency.

SANKHYĀ

The Indian Journal of Statistics

Edited by P. C. Mahalanobis

Vol. 4, Part 4, 1940

Articles on Theoretical Statistics

A Sample Survey of the Acreage under Jute in Bengal........P. C. Mahalanobis
Discussion on Planning of Experiments........................P. C. Mahalanobis
Technique of Random Sampling..................................K. B. Madhava
The Use and Distribution of the Studentized D^2-Statistic when the Variances and Covariances are based on K Samples.............S. N. Roy and R. C. Bose
The Median in Tests by Randomization..........................K. R. Nair
Table of Confidence Interval for the Median in Samples from any Continuous Population...K. R. Nair
A Study of Forty-three Years of Rainfall in Calcutta, 1893–1935......S. S. Bose
A Note on Examination Marks......................................Q. M. Hussain

Articles in the following fields are also included:

Economic and Business Statistics
Agricultural Statistics
Anthropology

Annual subscription rate: 20 rupees (or 32 shillings)
Inquiries and orders may be addressed to the Editor,
Sankhyā, Presidency College, Calcutta, India.

THE INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

OFFICERS FOR 1940

President:

S. S. Wilks, Princeton University, Princeton, N. J.

Vice-Presidents:

C. C. Craig, University of Michigan, Ann Arbor, Michigan.
A. T. Craig, University of Iowa, Iowa City, Iowa.

Secretary-Treasurer:

P. R. Rider, Washington University, St. Louis, Mo.

The purpose of the Institute of Mathematical Statistics is to stimulate research in the mathematical theory of statistics and to promote coöperation between the field of pure research and the fields of application.

Membership dues including subscription to the Annals of Mathematical Statistics are $5.00 per year. The dues and inquiries regarding membership in the Institute should be sent to the Secretary-Treasurer of the Institute.